Laura Verga: Learning Together or Learning Alone: Investigating the Role of Social Interaction in Second Language Word Learning. Leipzig: Max Planck Institute for Human Cognitive and Brain Sciences, 2015 (MPI Series in Human Cognitive and Brain Sciences; 170)

Learning together or learning alone:
Investigating the role of social interaction in second language word learning

Impressum

Max Planck Institute for Human Cognitive and Brain Sciences, 2015

Druck: Sächsisches Druck- und Verlagshaus Direct World, Dresden

Titelbild: © Laura Verga

ISBN 978-3-941504-55-4

Learning together or learning alone: Investigating the role of social interaction in second language word learning

Der Fakultät für Biowissenschaften, Pharmazie und Psychologie
der Universität Leipzig
eingereichte

DISSERTATION

zur Erlangung des akademischen Grades

doctor rerum naturalium
Dr. Rer. Nat.

vorgelegt von
Laura Verga, M.Sc.
geboren am 22 Februar 1985 in Busto Arsizio (Italien)

Leipzig, den 06.02.2015

Acknowledgments

“*Would it save you a lot of time if I just gave up and went mad now?*”

Douglas Adams, from “The Hitchhiker’s Guide to the Galaxy”

The blank page. My oh my, such a nightmare for every writer who ever lived! Lucky for me, at least in this part of the thesis it was not difficult at all to start, as it is really simple to know who to thank after these years.

First and foremost, I would like to thank my supervisor, Sonja Kotz, for her constant support and her capacity to be not only a great teacher but also an amazingly understanding person. For letting me explore my interests and support my intuitions and decisions, for being always available not only for work-related problems but also for everything else. I really learned a lot in these last years – as a student but even more as a person – and I am extremely grateful for this.

Second, I would like to thank the EBRAMUS group, for giving me the chance of pursuing this PhD in an international environment and for all the occasions we had to meet and discuss our work. I was lucky enough to meet wonderful people in this group, so my thank you goes to all the supervisors and students. In particular, I would like to thank Prof. Emmanuel Bigand, with whom I enjoyed a very useful month in Dijon as my secondment. I would also like to thank Philippe Lalitte, for helping me out to produce the music stimuli we used in our experiment.

Further, I would like to thank my group at the MPI. The beginning in a new country would have been much more difficult without such a supporting group, with so many kind people who happen to be also amazing scientists. Not only the constant support during the group meetings greatly helped me to advance in my PhD, but also the fun moments we shared at the office and outside significantly improved my German experience. My thoughts are going especially to Kathrin, Patri, Michael, Artyom and Ellen.

I am especially grateful I did my PhD at the MPI for all its facilities and incredibly prepared technical staff. In particular I would like to thank the MRI MTAs, Simone, Mandy and Anke, for their effort to understand my awful German and their kindness and professionalism. Further, a great thank you to Ina Koch, Kerstin Flake and Jan Bergmann.

Of course, as everyone knows, a PhD is not only work, but through work you get to know amazing people with whom to share the highs and the lows. For being such a great safety net I would like to thank the friends I met during these years: in particular, Laura and Iris. A very heart-felted thank you goes of course to Anna, Marieke, Charles, Floris, Tatiana, Irene, Luca, David, Christian, Franzi and Ellie (not in order of importance, nor alphabetical. I mixed you up randomly, ah! - No, seriously though. I created a random array in matlab). It would have most definitely not been the same without you guys. Also: Anna, your own dissertation was a great inspiration and life-saving example.

Last but not least, the two most important “thank you” of all. First, to my family, for always supporting me without questioning during these years and during the years before.

Finally, to Riccardo, without whom literally none of this would have been even remotely possible.

Table of contents

Preface

> *"No one can whistle a symphony. It takes a whole orchestra to play it."*
>
> Halford E. Luccock

From our very first day of life we are never truly alone. The social dimension represents one of the most relevant aspects of life. While the presence of another person is, in most cases, sought for its pleasantness, there are several instances in which social interaction becomes essential. For example, a knowledgeable partner is often required to acquire skills and knowledge necessary for life. Humans share this ability of learning from conspecifics with many other species (Frith & Frith, 2012; Hari & Kujala, 2009; Kuhl, 2007); however, this process becomes particularly evolved in infants when they learn the very complex code they will later on use to communicate: Language.

The case of language learning in infancy represents indeed a unique example of the relationship between learning and social interaction. In this context, social interaction is not only helpful for the learner, but it is mostly *necessary* for many aspects of language acquisition. For example, American infants were able to learn different Mandarin phonemes when learning from a real person, but not from a tape-recording (Kuhl, Tsao, & Liu, 2003). An explanation proposed for this phenomenon is that the caregiver represents a complex multimodal source of cues that may help the learner to pinpoint relevant elements in the environment (Waxman & Gelman, 2009) by promptly and adaptively coordinating to his behavior (Kuhl, 2007; Pereira, Smith, & Yu, 2008). In this way, the adult caregiver appears to the child as a particularly salient cue worth giving attention to. Accordingly, most models of first language learning emphasize the role of a caregiver in the learning process (Hollich et al., 2000; Kuhl, 2007; Tomasello, 2000).

Does this special link between language learning and social interaction end in adulthood? Are human adults self-sufficient learners? And how exactly should social interaction support adult word learning? Answers to these questions critically depend on a reflection on similarities and differences between first and second language learning (Rodríguez-Fornells, Cunillera, Mestres-Missé, & de Diego-Balaguer, 2009): If the processes underlying word acquisition support learning in both infant and adult learners, then social interaction should be considered as a potential factor even in second language learning. Instead, studies on word acquisition in adults have so far failed to investigate this aspect. Indeed, while the last decade saw an exponential increase of interest within the neurosciences toward social cognition, only a few studies have explored the impact of social interaction in complex communicative situations involving the use of language (Jeong et al., 2010, 2011). These studies have consistently shown how interaction with a social partner boosts brain activations in those brain areas which are usually recruited for the task at hand, such as areas involved in word learning in a verbal learning task (Jeong et al., 2010), or areas involved in communication during an interactive language interview (Jeong et al., 2011), thus suggesting an effective influence of social interaction during the task. However, several questions remain open: How exactly does this modulation occur? What are the basic mechanisms supporting social learning? Does social interaction exert a direct influence on the activity of high cognitive brain areas? Or does it modulate the activity of broad cognitive networks?

The aim of this dissertation is to deepen our knowledge of the mechanisms supporting word learning in healthy adults during social interaction. To this end, a game-like setting was developed to allow healthy young adults to learn new words interactively with another person. This set-up was adapted and employed to collect both behavioral as well as neuroimaging data.

Chapter 1[1] describes the theoretical background, the empirical issues and the open questions underlying the study of word learning during social interaction, together with the most relevant theories on word learning in a first and second language. In the first part of the chapter, the rationale for taking an interest in social cognition is outlined, together with the most recent advancements in this field of research. More specifically, the chapter highlights and explains what social interaction is, what its neural correlates are, and which methods can be used to study it. Furthermore, a possible role for social interaction during language learning is outlined in both children and adults. The second part of the chapter develops this theme, by pointing out not only *how* social interaction may be relevant for word learning, but - more specifically - *why* this should be the case. An analysis of the processes supporting language acquisition will highlight the similarities between learning mechanisms in first and second language learning. Despite this consistence, social interaction has until recently not been deemed influential for adults, who are often considered as self-sufficient learners. Recent evidence challenges this perspective, by showing that the context of language acquisition and usage may significantly shape word-learning processes, thus opening the possibility that adult learners may not be entirely self-sufficient after all.

Chapter 2 discusses the specific research question addressed in this dissertation and describes how each of the performed studies links back to the main theoretical framework, showing the continuity between them and their rationale.

Chapter 3 deals with the description of a new method that has been created ex-novo and employed in our experiments. The chapter further illustrates the

---

[1] The first part of this chapter has been published in a peer-reviewed journal (Verga & Kotz, 2013).

behavioral and neuroimaging methods and analyses used in the studies to provide a thorough characterization of the learning process during social interaction and its neural correlates.

Chapter 4 describes five studies[2] that were designed to validate the proposed method, and to subsequently investigate the behavioral and brain mechanisms supporting learning of new words during social interaction. The first two studies provide validation data for the set-up. Each of the subsequent studies focuses on a specific empirical question: The third study explored how social interaction influences participants' behavior during a word learning task with behavioral measures. Results of this study show that, during social interaction, temporal coordination emerges between the learner and the experimenter, which may serve to direct attention towards relevant elements in the learning environment. This result provides the first evidence that social interaction may serve word learning, by capitalizing on attentional mechanisms similarly to what has been proposed for first language learning.

The fourth study investigated whether the results of the first study are specific to social interaction, or whether a different stimulus with similar properties (in other words, being complex, rhythmic, and potentially interpersonally coordinated – M. Wilson and Wilson, 2005) may also elicit the development of a temporally coordinated behavior in the learner and, as a consequence, facilitate learning. Music represents a stimulus with these specified characteristics, and was chosen in this study as a comparison to investigate the specificity of social interaction in modulating word learning. The results of this study show that, despite music being able to stimulate temporal coordination, only social interaction has an impact on word learning. This outcome suggests that a social partner may represent a type of external stimulus particularly able to influence learning processes.

---

[2] Studies 3, 4 and 5 are based on manuscripts in preparation.

The final study investigated the neural basis of contextual word learning during social interaction (or a “second person approach”, Schilbach et al., 2013; Schilbach, 2014). This experimental question was tackled from several angles: First, a standard mass-univariate General Linear Model (GLM) analysis was conducted to elucidate which brain areas are specifically involved in our task; second, to investigate the brain networks involved in social learning, an Independent Component Analysis (ICA) was performed. This analysis was complemented by performing Dynamic Causal Modelling (DCM) of one of the relevant networks, namely the fronto-parietal attention-reorienting network. The results of this study suggest that social interaction significantly modulates task-related areas and their connectivity.

Finally, chapter 6 summarizes and discusses the results of all studies. It highlights the continuity between all of the studies, and emphasizes the feasibility of the developed setting as a promising tool to investigate social word learning. Furthermore, a novel model of social word learning in adults is proposed, and open questions and future directions are described in a critical overview. It is highlighted how the role played by social interaction may be critically considered as an influent variable whenever an interaction is present; in particular, this includes clinical approaches such as therapeutic settings for the treatment of language-related disorders such as aphasia.

1

Introduction

1.1 How relevant is social interaction in second language learning?[3]

In his book “Pragmatics of Human Communication” (Watzlawick, Bavelas, & Jackson, 1967), the psychologist and philosopher Paul Watzlawic stated that it is impossible not to communicate. Indeed, in his view, every behavior is a form of communication intended to convey a message from a sender to a receiver (Shannon & Weaver, 1963). The interaction between partners crucially defines a communicative intention: While a sender and a receiver do not have to be present at the same time or in the same place, there is no communication without one of the two partners. The interactive nature of this process is reflected in the word “communication”, meaning “share with someone”, “let someone know” (from the Latin cum – with - and munire - to bind together). However, the study of the most widespread vehicle of human communication, language, has so far suffered from an individualistic approach. Here, we review recent findings bridging social cognition and communication by highlighting

---

3 This chapter is based on the review article: Verga, L. & Kotz, S.A. (2013) How relevant is social interaction in second language learning? Frontiers in Human Neuroscience, 7:550, doi: 10.3389/fnhum.2013.00550

evidence that points towards the necessity to consider the impact of social interaction when investigating second language learning.[4]

1.1.1 Human communication and the role of social interaction

Human language is one of the most complex codes used to communicate between individuals. In its verbal form it is based on a small subset of sounds that can be combined in a potentially infinite number of bigger elements (words, phrases and sentences). The complexity of this code is further increased by the fact that human communication entails much more than the simple coding or decoding of linguistic utterances: For a communicative act to be effective, it is necessary for both the sender and receiver to understand the intentional state of a partner (De Ruiter et al., 2010; S. E. Newman-Norlund et al., 2009), an ability termed Theory of Mind (ToM) or mentalizing (Frith & Frith, 2006). The processes subtending ToM can be triggered by different contextual cues as long as they come from an agent (Frith & Frith, 2006); their function is to facilitate predictions about others' behavior via both verbal (Carruthers, 2002) and non-verbal (Noordzij et al., 2009; Willems et al., 2010) communication. An example of the latter case is reported in severe aphasic patients: Although virtually unable to express themselves verbally, these patients are able to pass tests intended to specifically tackle their residual communicative abilities; for example, they are able to engage in intention recognition with a partner in a non-verbal game requiring the patient to signal the position of a specific target on a checkerboard (Willems, Benn, Hagoort, Toni, & Varley, 2011; Willems & Varley, 2010). Another example comes from normally developing infants: Although they have not yet developed verbal language, they are able to use the caregiver’s gaze direction as a cue to orient attention; this behavior requires a proto-mentalizing ability to infer the caregiver’s intention and represents one of the first communicative acts in

---

[4] In the article Verga & Kotz (2013), this paragraph is entitled “Introduction”.

children (Csibra & Gergely, 2009; Tomasello & Carpenter, 2007; Tomasello, 1995; see below). In adults, mentalizing processes are activated by cues such as the identity of the person they are interacting with. In a recent study, Newmann-Norlund and colleagues demonstrated that in a non-verbal communicative task, adult participants adapted their communicative behavior to the presumed cognitive abilities of the partner. During the task, participants had to communicate to a partner the spatial location of a target on a checkerboard by moving a token to the position of the target, and they were told that the partner could either be an adult or a child. When they believed that they were interacting with a child, participants spent more time moving the cursor, thus emphasizing a crucial element of communication such as the target location (S. E. Newman-Norlund et al., 2009). When the partner is a peer, adults still adapt their behavior and, in most cases, this adaptation is reciprocal and results in behavioral resemblance between the partners. For example, pairs of adults tend to coordinate their body postures and gaze patterns during conversation even without being aware of it (Shockley, Baker, Richardson, & Fowler, 2007; Shockley, Richardson, & Dale, 2009), and also reduce the variability of their actions to better synchronize with each other (Vesper, van der Wel, Knoblich, & Sebanz, 2011, 2012). Another example is the tendency to share the feelings and emotions of others, often leading to the mimicry of an observed emotion (De Vignemont & Singer, 2006; Singer, 2006). An immediate evolutionary advantage of these phenomena is to facilitate learning mechanisms based on observation and imitation (Frith & Frith, 2012). However, how do these coordinative and imitative phenomena influence language? First of all, effective communication is based on the ability to know when it is the right moment to speak. This turn-taking ability relies on general coordinative rules, both on the side of motor coordination (Shockley et al., 2009) and on the side of conversation. For example, you do not want your partner to wait forever for an answer, but you also do not want to speak while he is still speaking ("minimal gap, minimal overlap" rule, Stivers et al., 2009). Furthermore, aspects of a conversation, such as the speaking rate and the similarity of words spoken in a dyad, also influence the coordinative pattern as demonstrated by Shockley and colleagues (2007): The authors showed that pairs of participants were maximally synchronized in their bodily movements when they were uttering the same words at the same time (Shockley et al., 2007). Even more importantly, imitative motor phenomena are influenced by the conceptual level of the conversation; for example, hand gestures in a conversation are likely to be imitated and repeated by the partners, but only if they make sense in the context of the speech (Mol, Krahmer, Maes, & Swerts, 2012).

Taken together, this evidence suggests that there is a two-way influence between social interaction and communication. However, the role played by social interaction has been greatly undervalued so far, especially in studies on language learning, even though this context represents a prototypical interactive communicative situation. In the following sections, we will first describe technical limitations that may have been responsible for such paucity in research; then, we highlight evidence on the impact of social interaction on learning in clinical and non-clinical populations.

1.1.2 Brain imaging in interacting individuals: Issues and solutions

One possible reason why social interaction has not, until recently, been considered as a factor in language learning studies, is the limitation that dual settings pose to imaging set-ups. Luckily, the influence of an interactive social approach has increased exponentially over the last decade (Galantucci & Sebanz, 2009; Knoblich & Sebanz, 2006, 2008; Schilbach et al., 2013), leading to an attempt to find new techniques and to create experimental situations tailored towards real-life situations often involving more than one person (Hasson, Ghazanfar, Galantucci, Garrod, & Keysers, 2012; Montague et al., 2002). This effort has led to the development of paradigms specifically intended to tackle social situations (Anders, Heinzle, Weiskopf, Ethofer, & Haynes, 2011; Schippers, Roebroeck, Renken, Nanetti, & Keysers, 2010), in which participants are often made to believe that they are interacting with someone. For example, pairs of participants may be required to take turns in the fMRI (functional Magnetic Resonance Imaging) scanner while observing a video recording of the partner during meaningful gestural (Redcay et al., 2010; Schippers, Gazzola, Goebel, & Keysers, 2009; Schippers et al., 2010) or affective (Anders et al., 2011) communication, while they believe this interaction is happening in real time. These kinds of “fake” communicative situations have allowed researchers to observe in-vivo activations in brain areas involved in the ToM system. This is supported by a network encompassing the medial prefrontal cortex (mPFC), the posterior superior temporal sulcus (pSTS), the temporo-parietal junction (TPJ), and the temporal poles (TP) (Amodio & Frith, 2006; Decety & Lamm, 2007; Frith & Frith, 2006; R. D. Newman-Norlund, Noordzij, Meulenbroek, & Bekkering, 2007; Noordzij et al., 2009; Saxe, 2006). Another system usually involved in “social” tasks is the human Mirror Neuron System (MNS). This system encompasses a fronto-parietal network of the ventral premotor cortex (vPMC), the inferior frontal gyrus (IFG), and the inferior parietal lobule in its rostral portion (iPL) (Rizzolatti & Craighero, 2004), and possibly other regions, including the dorsal premotor cortex (dPMC), the supplementary motor cortex (SMA), and the temporal lobe (Keysers & Gazzola, 2009). Important for the topic of this review, these ‘mirror’ neurons deal with the decoding of an action goal not only when one is performing an action, but also when observing the same action being performed by someone else (Keysers & Gazzola, 2009; Rizzolatti & Fabbri-Destro, 2008). These neurons thus provide an interface between one’s own motor repertoires and those of others (Knoblich & Sebanz, 2006). This “goal-sharing” property supports the hypothesis that brain areas exhibiting mirror-like properties should be more active during joint actions than during solitary actions (R. D. Newman-Norlund et al., 2007). Although ‘fake’ social interactive tasks allow this hypothesis to be indirectly tested, recent developments in neuroimaging have allowed the creation of new techniques to be applied to fMRI (Montague et al., 2002), Electroencephalography (EEG; Astolfi et al., 2010, 2011), and Near-Infrared Spectroscopy (NIRS; Cui, Bryant, & Reiss, 2012), enabling two (and sometimes more) people to be tested at the same time. These “hyper-scanning” techniques (Dumas, Lachat, Martinerie, Nadel, & George, 2011) allow ecologically valid interactions to be studied in a number of tasks, which could then also be applied to interactive learning paradigms. The clear advantage is that they allow a direct comparison of processes happening in two brains at the same time, a comparison that could otherwise only be inferred. Thus, one could potentially observe both the effects of mentalizing (Astolfi et al., 2010; Cui et al., 2012; King-Casas et al., 2005; Saito et al., 2010) and synchronization (Schippers et al., 2009; Tognoli, Lagarde, DeGuzman, & Kelso, 2007) on brain activity, in a real-time learning set-up. The use of hyper-scanning in these tasks demonstrates that not only is the behavior of two interacting people influenced by social interaction, but also their brain activation patterns. Indeed, synchronized EEG activity in frontal and central regions has been found in theta and delta oscillations of pairs of guitarists playing a melody together (Lindenberger, Li, Gruber, & Müller, 2009); similarly, when pairs of participants are required to spontaneously imitate each others, their brain activity becomes synchronized in the alpha-mu band over right-centro-parietal regions (Dumas, Nadel, Soussignan, Martinerie, & Garnero, 2010). Activity in this frequency band has been proposed to represent a neuromarker of human social coordination and, more specifically, has been linked to the human MNS (Tognoli et al., 2007). Saito and colleagues (2010) used fMRI hyper-scanning to scan two people at the same time while they were engaged in a real-time gaze exchange; that is, the pair were asked to direct one another’s attention to an object via eye movements. The authors found that the exchange of attention via eye gaze resulted in an inter-subject synchronization of the neural activity in the right IFG (Saito et al., 2010). Mentalizing and mirror systems thus seem to be recruited in social tasks (Ciaramidaro, Becchio, Colle, Bara, & Walter, 2014; Uddin, Iacoboni, Lange, & Keenan, 2007; Van Overwalle & Baetens, 2009), but their activity is influenced by the presence of a partner. Thus, the question arises: What happens in the case of learning a new language? A first attempt to answer this question arises from a recent study by Jeong and colleagues, in which the authors suggest that when words in a novel language are learnt in a social situation (but not when they are learnt from a text), the elicited brain activity (in the right supramarginal gyrus, SMG) is similar to the activity elicited by words in one's mother tongue (Jeong et al., 2010). However, the social situation depicted in this study was represented by movie clips of a dialogue. Thus, the question remains: What happens in a natural (social) learning situation?

1.1.3 Language learning and social interaction in children

As previously discussed, the ability to socially interact emerges very early in life (Grossmann & Johnson, 2007), and is represented by a number of basic interactions that children in the first year of life are able to master, such as following a caregiver's gaze, attracting her/his attention, and responding to her/his attentional requests. This set of abilities is usually grouped under the name "joint attention", entailing an interaction between a child, the caregiver, and the focus of attention (an object) (M. Carpenter, Nagell, Tomasello, Butterworth, & Moore, 1998; Mundy & Jarrold, 2010; Mundy & Newell, 2007; Mundy & Sigman, 2006). From a psychological point of view, the role of triadic attention ability during childhood is to create a common psychological ground shared between the infant and the caregiver, and relies on the formation of ToM in children (Tomasello, 1995). In this common space, adults act as experts and guide the children toward the relevant information that should be learnt, by using an effective signal such as eye gaze (De Jaegher, Di Paolo, &

Gallagher, 2010; Csibra & Gergely, 2009). In this asymmetrical learning setting, children's behavior is further facilitated by the fact that adults tend to adapt their communicative behavior by emphasizing crucial aspects of communication (for example, by spending more time on them) (S.E. Newman-Norlund et al., 2009). Moreover, the interaction with the caregiver increases motivation, thus reinforcing a given behavior (Hari & Kujala, 2009; Syal & Finlay, 2011; Vrtička, Andersson, Grandjean, Sander, & Vuilleumier, 2008). This asymmetrical learning setting, in which knowledge is passed from parents to offspring, is not limited to humans and can be found, for example, in many bird species that use complex vocal codes to communicate (Hari & Kujala, 2009; Frith & Frith, 2012; Kuhl, 2007). However, ToM abilities underlying human communication seem to represent a unicum in nature. Indeed, even our closer animal relatives, the chimpanzees, do not have the human ability to really "share" intentionality. As an example, chimpanzees are perfectly able to follow the gaze of an interacting human, but they do not try to start joint attention, nor do they try to infer the referent of the gaze as human children do (Tomasello & Carpenter, 2007). This human ability to share intentionality and acquired knowledge with other humans has been proposed to be at the core of the evolution of verbal language (Pinker, 2010; Tomasello, 1995). A series of experiments conducted by Kuhl and colleagues aimed to investigate this possibility and to test the impact of social interaction on phonetic discrimination in children (Kuhl et al., 2003; Kuhl, 2007). Cohorts of American infants were exposed to native speakers of Mandarin Chinese either via direct interaction or via pre-recorded videotapes, and subsequently performed a phonetic discrimination task. Interestingly, infants were able to learn different Mandarin phonemes when they were exposed to them via a real person, but not when the exposure was merely via a recording (Kuhl et al., 2003). There are two plausible explanations for this effect; first, a live human may attract more attention and increase motivation, as compared to a recording. Second, a real person can provide referential information, crucial for linking words and concepts (Waxman & Gelman, 2009). In particular, Kuhl and colleagues pointed out that joint attention towards an object being named can facilitate a child's capacity for word segmentation (Kuhl et al., 2003). Similarly, results from Hirotani and colleagues (2009) suggest that joint attention helps to strengthen the association between a word and its referent, thus facilitating learning. These authors found that semantic integration, reflected in the N400 effect, seemed to be present when children learnt new words in a joint attention condition but not in a non-joint-attention context. Although infant learning represents a particular case, vocabulary learning poses similar demands to both children learning their first language, and to adults when learning a new language. Thus, factors facilitating word learning in children could potentially impact adult learners in a similar way.

1.1.4 The role of sociality in second language learning

Evidence thus accumulates to favor the view that the development of verbal language is, at least, supported by establishing common ground between a sender and a receiver. In turn, the events that take place in such common space are mostly dependent on the interaction between partners (Mundy & Jarrold, 2010). However, a note of caution needs to be used when comparing language learning in children and in adults. Indeed, learning of a second language (L2) can occur largely independently of the presence of another person, and is usually learnt via explicit formal training as compared to a first language (L1), which is acquired effortlessly and without explicit instructions (Abutalebi, 2008). Nevertheless, the case of word learning represents a link between language learning in infants and in adults. Indeed, words in a new language can be acquired incidentally (Nagy, Anderson, & Herman, 1987; Swanborn & De Glopper, 1999; Laufer & Hulstijn, 2001; Rodríguez-Fornells et al., 2009), for example, new words encountered while reading a text can be easily learnt. In this situation, an adult learner faces the same problems as an infant, namely the indetermination of the referents (that is, there are multiple words in a language and multiple possible referents in terms of meaning). However, how can the correct meaning be assigned to an unknown word? The easiest way to go about this problem is exemplified by associative learning, a procedure that concentrates on the statistical learning of the co-occurrence of data from speech and its context (Breitenstein, Kamping, Jansen, Schomacher, & Knecht, 2004; Whiting, Chenery, Chalk, Darnell, & Copland, 2007, 2008). The advantage of this procedure is that it poses low cognitive demands during training (Dobel et al., 2010; Pulvermüller, 1999) and is resistant to errors made during a phase of guessing (S. K. Carpenter, Sachs, Martin, Schmidt, & Looft, 2011). The underlying rationale is that once a word is heard in an utterance or seen in a sentence, a set of potential meanings can be inferred from the context, thus reducing the number of possible referents (Adelman, Brown, & Quesada, 2006). This way, novel word forms can be acquired and integrated in the lexicon relatively quickly and successfully. For instance, neural responses evoked after training are indistinguishable from those obtained in response to "old" words, as demonstrated in the disappearance or reduction of an N400 response (Mestres-Missé, Rodriguez-Fornells, & Münte, 2007). The N400 component is a negative deflection starting 200-300 ms after the presentation of a word, and has been associated with semantic processing (Lau, Phillips, & Poeppel, 2008). Its disappearance in a learning paradigm thus possibly corresponds to establishing a link between a novel lexeme and conceptual information (Dobel et al., 2010; Mestres-Missé et al., 2007). The neural network supporting word learning involves regions of the semantic circuitry such as the left IFG (BA45), the MTG (BA21), the parahippocampal gyrus, and several subcortical structures (Mestres-Missé, Càmara, Rodriguez-Fornells, Rotte, & Münte, 2008). Although, in adults, new vocabulary can be learnt independently of the presence of a partner, social interaction may increase the number of cues and referential information in much the same way as it does in infant learning (Kuhl, 2004, 2007, 2010). Indeed, the interaction between partners in conversation could lead L2 learners to focus on certain aspects of the context and certain words in speech (Yu & Ballard, 2007). The coordinative phenomena we describe above could play a role in this process, maximizing the efficiency of the conversation and consequently facilitating the focusing of attention. This proposal has been made for word learning in toddlers; indeed, it has been shown that in toddler-adult dyads, the number of new words learnt by toddlers is proportional to the quality of the synchronization during the interaction with the caregiver (Pereira et al., 2008). Again, it is important to note that the case of word learning is not dissimilar in adults and infants, and so one may expect facilitating factors (such as the focusing of attention driven by synchronization) to play a role in word learning for both adults and children. Indeed, although it is possible to learn a new language alone, adults often learn a new language in social contexts, most commonly in a teacher-learner setting, which requires interaction with a partner as well as sophisticated reading of a speaker's intentions (P. Bloom, 2002; Mestres-Missé et al., 2007; Mestres-Missé, Münte, & Rodriguez-Fornells, 2009). Thus, the necessity to consider sociality as a factor in L2 studies seems striking, particularly in light of evidence that when new words are encoded in a social context, but not when they are learnt by translation, the pattern of activation in the retrieval phase is similar to the one observed for L1 words (Jeong et al., 2010).

1.1.5 Learning and social cognition in pathological populations

Learning new words, or re-learning words that have been forgotten, is the goal not only for infants and L2 learners, but also for pathological populations including, for example, people suffering from autism, dementia, or aphasia. In these pathological populations, the role of social interaction is becoming increasingly acknowledged as a crucial variable for therapeutic outcome success. Communicative deficits in autism spectrum disorders have been frequently attributed to higher cognitive processing impairments, and especially to ToM deficits (Baron-Cohen, Leslie, & Frith, 1985). However, more recent evidence indicates that lower-level processes may also be affected. For instance, recent findings suggest that autistic children display low-level difficulties in temporal processing, including impaired timing and deficits in the perceived duration of an event, which can in turn influence the perception of relevant social cues such as eye gaze (Allman, Pelphrey, & Meck, 2012; Allman, 2011; Falter, Noreika, Wearden, & Bailey, 2011; Falter & Noreika, 2011). The fact that ToM and timing abilities may be crucial for language, even in a population who display impaired ToM, comes from the discovery that autistic children improve their language abilities after a treatment focusing on the optimization of their joint attention capacities (Kasari, Paparella, Freeman, & Jahromi, 2008).

Similarly, social interaction plays a role in language re-learning in aphasia. A paradigmatic example of this claim comes from a specific form of therapy for severe aphasic patients based on music, namely Melodic Intonation Therapy (MIT; Norton, Zipse, Marchina, & Schlaug, 2009). This approach uses musical and sensory stimulation in order to improve the speech production of the aphasic patient and is centered on the role of the therapist. Although the beneficial effect of the therapy has been traditionally attributed to the effect of music tout court, recent evidence challenges this perspective and suggests that rhythm (and not necessarily melody) holds the key to understanding the impact of music therapy (Stahl, Kotz, Henseler, Turner, & Geyer, 2011). Considering that music therapy is therapist-centered, this result fits a joint-action explanation well, as rhythm is defined by the coordinated action between a therapist and a patient. This coordination strongly influences timing and its variability of the single individual in the interaction. Future investigations should attempt to differentiate the role played by joint action dynamics from those played by the timing of the interaction per se.

1.1.6 Summary[5]

In conclusion, the role of social interaction in language learning has, thus far, been widely overlooked, partly because of the technical constraints posed by interactive settings in imaging studies. We propose that further studies on language learning in adults should further explore the powerful impact of social interaction. This necessity comes from at least four lines of research: First, language use intended as communication is an interactive phenomenon, relying on the ability of partners to infer each other's mental state and to coordinate with each other in successful turn-taking. Second, in infants, joint attention with a caregiver provides additional contextual cues that drive attention and motivation, and that can help to disambiguate the meaning of a new word (or stimulus); analogously, contextual learning represents one of the easiest ways for late learners to acquire new words and can thus be influenced in a similar way by social interaction. Third, and related to the second, the investigation of interactive language learning resembles a natural learning situation involving a teacher and a student. Fourth, the role of sociality is starting to emerge as a valid explanatory variable in the context of word learning in pathological populations.

---

5 In the article Verga & Kotz (2013), this paragraph is titled "Concluding remarks".

1.2 Word learning in first and second language

The previous chapter posed the question of whether social interaction is relevant to adult learners of a second language. This query is motivated by the observation that language learning has, so far, mostly been studied with an individualistic approach. This perspective seems particularly odd – or, at least, incomplete - considering that, for humans, language represents the most relevant vehicle of communication; as such, it is used to convey information *between* people. Despite the methodological complications that a dyadic setting may pose, it is then critical to ask the question of whether social interaction should be included in the variables influencing learning of both first and second languages.

Nevertheless, this question is only the tip of the iceberg of a set of open, corollary issues; in order to provide an answer to the original question (*is social interaction a variable influencing word learning?*) it is first necessary to understand which mechanisms support word learning, which ones among them are modulated by external inputs and, finally, whether the impact of an external input is independent of the experience of the learner: Naïve learners might be more dependent on outside information, whereas experienced learners might be better equipped to efficiently integrate multiple sources of information. This latter point corresponds to two different but related questions: First, whether first and second language learners rely on the same mechanisms to acquire new words; second, whether external inputs are weighted and used differently depending on the expertise of the learner. These questions remain open, as the processes and mechanisms underlying word learning in children and adults have rarely been compared. However, similarities and differences in the way the two populations acquire a language may be particularly helpful in understanding the learning mechanisms required to master a language (Rodríguez-Fornells et al., 2009).

The present chapter provides a more in-depth analysis of the processes supporting word learning in both children and adults. Further, the most influential theories of word learning are described, followed by the differences and commonalities between first and second language acquisition. The aim of this chapter is therefore to provide a theoretical backbone, grounded in the research on word learning, with the goal of proposing not only *how* social interaction might be relevant to word learning but, most notably, *why*.

1.2.1 Processes of word learning

> "*Before children can tie their shoes, they have mastered thousands of words.*"
>
> George J. Hollic

Children are remarkable learners. A newborn baby is completely dependent on her/his caregiver to survive; yet, the baby is able to learn an unparalleled amount of information in a relatively short period of time. The effort of the caregiver in raising the child is not unrewarded, as the very first words the infant pronounces (usually, "mama" or "dadda") are the names of those who raised her/him so thoughtfully. This moment not only represents a very emotional event for the caregivers, but also signals a crucial cornerstone for the child's development. Indeed, these first verbal expressions, appearing around one year of age, start a lifelong and constantly improving process becoming faster and more efficient day by day. Soon after this moment the child's vocabulary will encounter an exponential growth, with an average number of 10 new words acquired every day (P. Bloom, 2001; Carey, 1978; Hollich et al., 2000). How is this possible? How do children *learn how to learn* in such an efficient manner? This evidence is extraordinary when considering infants' limited cognitive capacities, but does not become less astonishing when thinking about adult learners. Consider, for example, an adult travelling abroad in a country she/he does not understand the language of. In this scenario, the traveller is in much the same learning situation as an infant: The speech she/he is confronted with is a flow of uninterrupted sounds, which needs to be deconstructed into basic elements (that is, the words and their meanings) to be understood and used.

At first glance, the steps required to learn new words are remarkably simple for both infants and adults: First, learners need to separate the continuous flow of speech into specific units - the new words. Then, they need to attach these verbal labels to their corresponding meanings, and to identify superordinate and subordinate categories to which the same label applies. However, things are not as easy as they may seem, and every step in the process hides multiple issues with little hope of an easy solution. For example, how does a learner identify what a new word means in a world in which infinite meanings exist? This critical doubt suggests that the pitfalls of word learning are not to be underestimated, and brings us back to the original question: How is it possible to elude these difficulties? In the following, the answer to this question will be provided thorough a description of the processes and mechanisms supporting the initial two phases of word learning, namely *segmentation* and *referencing*.

1.2.1.1 Words and sounds: Speech segmentation

The chain of events leading a naïve learner to become a language expert is – theoretically – quite straightforward, and begins with the discovery that specific sequences of sounds go together to form word units. This would not be a difficult task if it were not for one crucial detail: Often, there are no easily identifiable blank spaces clearly marking the boundary of each word; instead, everything is presented in a mostly uninterrupted sound stream. Children are facilitated with the speech input, as caregivers tend to use a simplified type of speech input characterized by enhanced prosodic cues, including longer pauses (L. Gleitman, Newport, & Gleitman, 1984; Thiessen, Hill, & Saffran, 2005). Even with this facilitation, to separate the speech flow into specific and separate units – the new words – is not a trivial task. This first step in word learning is known as the *segmentation problem* (Brent, 1999).

How do children and adult learners solve this apparently impossible task? Research on speech segmentation provides abundant evidence that learners are able to break up the speech stream by using several different cues to identify words' boundaries. It is important to note that while these cues were originally discovered in children, they represent powerful mechanisms employed also by adults in word learning, as will be specified in the following paragraphs (McQueen, 1998; Norris, McQueen, Cutler, & Butterfield, 1997; Perruchet & Vinter, 1998; Saffran, Johnson, Aslin, & Newport, 1999).

Prosodic (or rhythmic) cues are employed to identify which letter strings correspond to word units (Jusczyk, Houston, & Newsome, 1999). This type of cue includes aspects such as the stress pattern of the mother tongue, which is typical and recognizable already by infants as young as 7.5 months (Jusczyk et al., 1999). For example, 9-month-old American children prefer listening to words that have a sound pattern corresponding to English than to Dutch, while the opposite pattern is found for Dutch infants (Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993). Similarly, adult learners use prosodic cues to identify word units in a nonsense speech stream (Cutler & Norris, 1988; Saffran, Newport, & Aslin, 1996) and to resolve lexical ambiguities (Christophe, Gout, Peperkamp, & Morgan, 2003; Shukla, Nespor, & Mehler, 2007).

The position of *allophones* (the same variant of different phonemes) is another cue used to identify a specific position in the word; for example, some phonemes are pronounced differently according to their position at the beginning compared to the end of a word (Christophe, Dupoux, Bertoncini, & Mehler, 1994; Church, 1987).

The role played by *phonotactic* rules is very well documented in both children (Mattys, Jusczyk, Luce, & Morgan, 1999) and adults (McQueen, 1998; Norris et al., 1997). In brief, each language is characterized by particular letter sequences which are allowed to occur within words, while other string sequences are not allowed. Since this latter type of string cannot be found within words, their presence in the speech flow most likely signals the transition between two different words.

Lastly, the probability of transition from one syllable to another – or *statistical cue* - is a powerful hint as to where the word boundary might be (Saffran, Aslin, & Newport, 1996; Thiessen & Saffran, 2003; François & Schön, 2013): Transitional probabilities are higher between syllables within a word and lower at the word boundary.

These cues provide first and second language learners with important information as to where the word boundaries might be, thus helping in breaking down the continuous stream of sounds into word units. Hence, at this stage, the learner is provided with some verbal labels; she/he knows each one is associated with a specific sound, and is able to track them even when presented continuously in fluent speech. Yet, another, and more crucial, question is to be answered: What do they mean?

1.2.1.2 Words and meanings: Indeterminacy of the referent and generalization

Words are symbols; this means that a word is – per se – nothing more than a convention expressed by a specific sound. What makes it special is that it stands for something else. However, how do learners identify what these “verbal labels” stand for in a world offering infinite possible meanings? This fundamental question – known as the *indeterminacy of the referent* - has no trivial answer. Willard Quine proposed an exemplary representation of this problem in the renowned "gavagai" example (Quine, 1960): An English linguist wanders in a foreign country where natives speak an unknown language, named Arunta. Suddenly, a rabbit passes by; a native points at it and exclaims: "*Gavagai!*". Most likely, the linguist will assume this new word to be the name of the rabbit. But does it really? How does the linguist know that "gavagai" refers to the entire rabbit? It might very well be just its tail, or its hopping. Is it the name of that specific rabbit, or is it the general label for all rabbits? And what if the native was not pointing at all? In this nightmarish scenario, the number of possible referents for "gavagai" in the visual scene is unlimited.

Despite these difficulties, both children (Clark, 1987) and adults (Yu & Smith, 2007; Yu, Smith, Klein, & Shiffrin, 2007) are perfectly able to ascertain the meaning of "gavagai", by tracking consistencies among different presentations of the same verbal label: If the Arunta speaker introduces his pet rabbit by saying "I have a Gavagai", the linguist will have some more information to conclude – for example - that "gavagai" does not mean "hopping". Additional information will be added if another Arunta native says "my Gavagai's name is Pip", and so on. This way, several exposures with the word will enable the linguist to map the new word to its meaning with reasonable certainty.

In most cases, however, it is not even necessary for a new word to be repeated several times; indeed, it has been shown that even young infants are able to "fast map" (Carey, 1978; Carey & Bartlett, 1978) a word onto its meaning, by using both linguistic and extra-linguistic cues to make first hypotheses on the alleged referent (Heibeck & Markman, 1987). For example, Carey and Barlett presented 3- and 4-year-old children with a new verbal label depicting a color, "chromium". In their study, children were asked to bring the experimenter "the chromium plate, not the red one". Results demonstrated that a single exposure to the new word was sufficient to induce some understanding of the new meaning, persistent even when tested a week later (Carey & Bartlett, 1978). Similarly, Mestres-Missé and colleagues (Mestres-Missé et al., 2007) demonstrated that adults are able to guess the meaning of a novel verbal label after just a few encounters with sentences containing the word. Likewise, Borovsky and colleagues (Borovsky, Kutas, & Elman, 2010) showed that adult learners have the capacity to correctly infer a word meaning even after a single exposure, when the sentence context limits the number of possible referents. When the sentence context is not restrictive enough towards a specific meaning, this first mapping is, of course, imperfect, and is then enriched and specified with further exposures to a word (Carey, 1978). In this case, the more diverse the information provided by the context, the more reliable the mapping should be between label and meaning (the importance of a varied sentence context is examined in more detail in paragraph 1.2.1.3).

Nevertheless, the fact that a reasonable mapping may be achieved so easily suggests that learners use strategies not only to limit the number of possible referents (*indeterminacy of the referent* problem), but also to understand which objects go together under the same label (*generalization* problem). Research conducted on this topic has identified several possible constraints that may guide the initial effort of the learner.

Representational constraints are a class of innate biases based on physical features of an object (Markman, 1990). To this category belongs the *whole object* assumption, which explains the evidence that both children (Markman, 1990, 1991; Waxman & Kosowski, 1990) and adults (L. Bloom, 2000) tend to interpret a new word as a label describing an entire object; to stay with Quine's tale, this assumption predicts that, when first hearing the Arunta speaker, learners would be biased to assume that "gavagai" means "rabbit". Of course, this bias is insufficient by itself to explain how words are learned as it leaves many open questions, such as how it is possible to learn words referring to parts of objects. To address this shortcoming another principle has been postulated, namely the *mutual exclusivity* principle (Markman, 1990, 1991; Markman & Wachtel, 1988; Merriman, Bowman, & MacWhinney, 1989): In brief, this bias predicts that if the learner knows that “gavagai” means rabbit, and someone points to it and says “Oh look! A *pliu*!”, then “*pliu*” will be assumed to refer either to a novel object or to a part of the rabbit. While Markman described these biases as innate predispositions (Markman, 1990), at least for the mutual exclusivity bias, a more intuitive explanation is possible: The learner simply does not accept two names for the same object. This assumption is known as the *principle of lexical contrast* (Carey & Bartlett, 1978; Clark, 1987; Nelson, 1988). Importantly, it has been demonstrated that adult learners also employ these principles. Indeed, by comparing word learning in children and adults, Golinkoff and colleagues (Golinkoff, Hirsh-Pasek, Bailey, & Wenger, 1992) showed that even adult learners use lexical principles, such as the mutual exclusivity bias, to solve the indeterminacy of the referent problem.

Another type of a-priori assumption, halfway between the two problems of reference and generalization, is the *taxonomic bias* (Markman, 1991). A study by Markman and Hutchinson demonstrated the existence of this assumption in preschool children, who were exposed to a series of objects. Each object was followed by two associate words, one having a thematic relationship with the target (for example, Dog – Bone) and one being in a categorical relationship (for example, Dog – Cat). The authors then asked children to choose an object similar to the target, described using a novel word (for example, “See this fep? Find another fep”). Results show that children consistently identified as “fep” objects belonging to the same taxonomic category as the original target (Markman & Hutchinson, 1984). Specific *linguistic* constraints have also been postulated as pivotal for word learning in both children and adults (Gillette, Gleitman, Gleitman, & Lederer, 1999; Landau, Smith, & Jones, 1992). These include primarily phonetic and prosodic information that facilitate not only segmentation, but also the word-meaning mapping; indeed, open-class words following a clear stress pattern (for example, nouns, verbs) are learned more easily and way before close-class unaccented words (for example, functors; L. Gleitman, Gleitman, Landau, & Wanner, 1988).

Learning how to map words to single objects does not represent the end of the learning journey; indeed, learners still need to understand that some labels may be applied to some objects, but not to others. In other words, they have yet to solve the *generalization* problem. Even in this case, a solution based on a-priori biases has been hypothesized. One very straightforward hypothesis is that both adults and children may have a specific *attention bias* toward particular properties of an object, such as its *shape* (Landau, Smith, & Jones, 1998; L. B. Smith, 2000); for example, if a novel object is presented and described as a “dax”, children will assume all objects with the same shape to be a “dax”. Adult learners still generalize a novel name by object shape, but also by object *function* (Landau et al., 1998), and it has been suggested that a developmental change may occur in the process of generalization, from a shape-based to a function-based process (Graham, Williams, & Huber, 1999). In both children and adults, the mechanism behind this phenomenon may be associative learning itself: Whenever one perceptual cue is regularly associated with another, the presence of the first will automatically increase attention to the second (A. Allport, 1989). However, in order for this control of selective attention to take place, several encounters with the new word are required to statistically identify commonalities among objects with the same shape or function (Landau et al., 1992).

The solutions proposed so far for both referencing and generalization focus on the properties of the visual or linguistic input: By being exposed to objects and speech, learners are able to track statistical recurrences and consistencies leading to an accurate word-concept mapping. So far, evidence supports the idea of largely overlapping processes supporting word learning in L1 and L2. But does everything depend upon the learner’s abilities? The answer to this question signals the first difference between word learning accounts for children and adults. In first language learning, several authors suggested that an external control mechanism may be involved in directing attention towards the referents, thus significantly reducing referential uncertainty (Baldwin, 1993; Dominey & Dodane, 2004; Gelman, 2009; Tomasello, 2000; Verga & Kotz, 2013). While the exact "external mechanism" may differ slightly from one author to another, they all agree that this mechanism may be defined by the label of *social information*. In this context, social information means that a social agent provides additional cues as to what the word means, where its boundaries are, or what it refers to. Classically, the label "social information" has been used in L1 models to identify two types of cues that may help the learner (Dominey & Dodane, 2004): First, *child directed speech* directs the attention of the child towards the relevant aspects in the sentence through modulation of the acoustic signal; second, the number of referents may be reduced by directing the child's attention to the appropriate aspect of the visual scene through a *joint attention* situation between child and caregiver. This hypothesis is supported by evidence of a positive correlation between joint attention abilities and the expansion of vocabulary (M. Carpenter et al., 1998) and, vice versa, between failures in joint attention and deficient linguistic abilities in autistic children (Mundy, 1995; Mundy, Sigman, & Kasari, 1990).

To return to the question previously formulated (that is, *does everything depend upon the learner's abilities?*), second language has been classically considered as a "learner-based" process; different to children, adult learners are not assumed to require social information in order to acquire a new language. However, this classical position has recently been challenged by evidence showing that socially grounded variables (for example, the level of exposure to a language) are crucial for language learning in adult speakers (see paragraph 1.2.3). The fact that social information may be crucial is a natural consequence of what was previously emphasized in this dissertation: First, language represents a form of communication with which to share information *between* people, who reciprocally influence each other in the process. Further, while child directed speech is, by definition, a phenomenon predominantly influencing L1 acquisition, joint attention is a social ability widely persisting and used in adulthood (Saito et al., 2010; Williams, Waiter, Perra, Perrett, & Whiten, 2005). Hence, the question arises as to whether a similar attentional account may be relevant to explain evidence of a social influence on adult word learning.

1.2.1.3 Remarks on word learning processes

In the previous paragraphs, the processes underlying the first steps in word learning have been outlined. These processes describe some relatively easy heuristics that both children and adults use to crack the speech code and identify the meaning of words, including statistical, phonotactic, syntactic, and social properties of the input, together with some interpretation biases. With regard to the latter, it is vital to point out that these heuristics represent – as the word suggests – *only a bias*, not a rule. In other words, they simply assert that upon hearing a novel word, the learner is prone to giving a particular interpretation, leading to the rapid creation of a temporary word-meaning link. This does not mean that this interpretation cannot be changed; instead, it will be successively specified by further encounters with the word, in which new elements from the sentence context will be extracted to refine the meaning of the verbal label. This "slower" mapping is critically dependent upon the variability in the information provided by the sentence context. While, in general, the importance of repetition in reinforcing the memory trace of the word-meaning pair has long been established[6], the role of the linguistic context of repetition is still a matter of debate (Besson & Kutas, 1993); indeed, results

---

[6] According to Besson and Kutas (1993) repetition priming may be defined as "the facilitation in processing accorded to a word, (which) is a consequence of the trace left by a previous encounter with a word".

have been found either pointing toward a repetition effect independent of the linguistic context (for example, Carr, Brown, & Charalambous, 1989) or, conversely, to a context specificity effect (for example, Den Heyer, 1986). To solve this incongruence, Besson and Kutas (1993) examined the effect of linguistic context on repetition by repeating the sentence context, the final word, both or neither. The authors hypothesized that if the linguistic context influences word repetition, then larger facilitation (described as better cued recall and a smaller N400) should be elicited by words repeated in the same context. Instead, if the word repetition effect is merely a function of the activation of a word's mental representation, then facilitation should be larger for repeated than unrepeated words, regardless of the context. The authors observed larger repetition effects for words repeated in consistent contexts, interpreted as evidence that word repetition is dependent upon episodic memory processes. In other words, linguistic context creates a specific memory trace of a unitary event. This evidence has important implications for word encoding: If a word repeated in a consistent context is encoded as a unitary episode, its mnemonic representation should be highly consistent, yet difficult to extend to novel contexts. As of yet, evidence supports different theories of human memory (Anderson & Bower, 1972; Waxman & Gelman, 2009), either showing a prevalence of more consistent context (Dempster, 1987; Hicks, Marsh, & Cook, 2005; Koffka, 2013; Steyvers & Malmberg, 2003; Young & Bellezza, 1982) or the importance of variability in the context as a successful mnemonic device (Hills, Maouene, Riordan, & Smith, 2010; L. B. Smith, 2000). In the latter case, variability is claimed to improve generalization to novel items in particular (Perry, Samuelson, Malloy, & Schiffer, 2010). This crucial point represents one of the research questions that has motivated the studies presented in this dissertation, and will be further explored in chapter 3.

One final, and crucial, point should be noted here: The description of learning heuristics for referencing and generalization has been divided – in line with the literature – between representational, linguistic and social processes. However, this distinction is artificial and misleading, as all the biases and constraints described are (or, at least, contain) social information. This evidence is often underestimated: Mechanisms facilitating learning are searched for in the input, as if the input could be produced by itself. Again, this claim is over-simplified; especially for children, there is simply no input without a human agent, and some types of information are not acquired if they do not come from a human agent (Kuhl et al., 2003). This type of evidence supports the hypothesis that word learning may represent a special instance of a more general capacity to acquire socially transmitted information (Csibra & Gergely, 2009; Markson & Bloom, 1997), which may not even be limited to humans, but may extend to other mammals such as dogs (P. Bloom, 2004; Kaminski, Call, & Fischer, 2004; Markman & Abelev, 2004)[7].

1.2.2 Models of word learning

“*Most people think it died with Behaviorism,*
but associative learning theory lives on.”
Michael Tomasello

Several theories have been developed to describe how words are learned (Hollich et al., 2000; Kuhl, 2007; Markman, 1991; L. B. Smith, 2000; Tomasello, 2000). Each model differs with respect to the weight given to specific assumptions and biases used by the learner during the process. For

---

[7] The study by Kaminski and colleagues investigated the fast-mapping ability of a domestic dog, a Border Collie named Rico. The authors not only found that Rico knew at least 200 words, but also that he was able to successfully learn and retrieve new words right away. This result was interpreted as evidence that “fast mapping appears to be mediated by general learning and memory mechanisms also found in other animals and not by a language acquisition device that is special to humans”. However, as pointed out by P. Bloom (2004), Rico was only learning during a specific fetching game, in which it was always the owner who was communicating with the dog. Further studies would be required to clarify whether Rico did in fact know the referent of the word, or whether he simply learned an object-name association.

example, the *associative theory* (L. B. Smith, 2000) emphasizes the role of basic attentional mechanisms in facilitating the association between word and meaning. The *constraint* (or *principle) theory* (Markman, 1990) puts the focus on a-priori biases leading to a preferred interpretation for new verbal labels, such as the whole-object bias. While these two theories emphasize the role of the learner in interpreting the input signal, the *social-pragmatic theory* (Tomasello, 2000) shifts the focus to communication between the learner and the caregiver; that is, learning a new language is substantially a matter of social interaction. Yet, despite their differences, these theories concentrate on specific and distinct aspects of the learning experience. The *emergentist coalition model* (Hollich et al., 2000) tries to integrate the tenets of these models into a unified and more complete theory of word learning. Even more comprehensive is the *social gating hypothesis* (Kuhl, 2007), stating that learners may be predisposed to pay particular attention to linguistic information when they are presented in a social context. This latter model proposes a learning hypothesis firstly investigated in, and applied to, language learning, but not limited to this specific field. In the following section, a more detailed explanation of the core tenets of these theories will be provided[8].

1.2.2.1 The associative theory

What difference exists between word learning, in which a label has to be associated with an object, and any other type of associative learning? None, according to Smith (L. B. Smith, 2000). In his view, word learning consists of

---

[8] Another theory of word learning is the syntactic bootstrapping hypothesis (L. Gleitman, 1990). This theory will not be described due to its linguistic nature, with a prevalent focus on the importance of syntax in verb learning. Instead, the other theories summarized here provide general psychological approaches focusing in general on words and, more specifically, nouns. The bootstrapping theory is mentioned in paragraph 1.2.1.2: In brief, this hypothesis suggests a mutual influence of syntactic and semantic information, whereby verbs are identified first and then help in interpreting the rest of the signal. Similarly, the bootstrapping prosodic hypothesis specifies that prosodic cues in speech may inform the child about the syntactic properties of the signal (L. Gleitman, Gleitman, Landau, & Wanner, 1988; Wanner & Gleitman, 1982).

associating a specific sound (the new word) with the most salient element of the contextual environment.

The term “salience” refers in this context to the ability of an object to stand out among other stimuli, often because its properties contrast with its neighbors. **Figure 1** represents a typical example of what a salient stimulus looks like in a visual scene: Considering the grey-scaled background, the colored hot air balloon catches the eye of the observer. A similar example in an auditory scene may be, for example, an abrupt laugh in a silent environment. In both cases, the effect of the salient stimulus is to attract bottom-up attention (for example, Parkhurst, Law, & Niebur, 2002). However, salience is not the only attribute of an object needed to create a link between two objects (or events). Indeed, in order for two events to become associated there must be a temporal contiguity between them, which means that they must occur very closely in time (Balsam, Drew, & Gallistel, 2010). *The associative theory* suggests that when a new word is presented, the learner immediately scans the environment is search for the most salient object, which is selected as the correct referent. In this view,

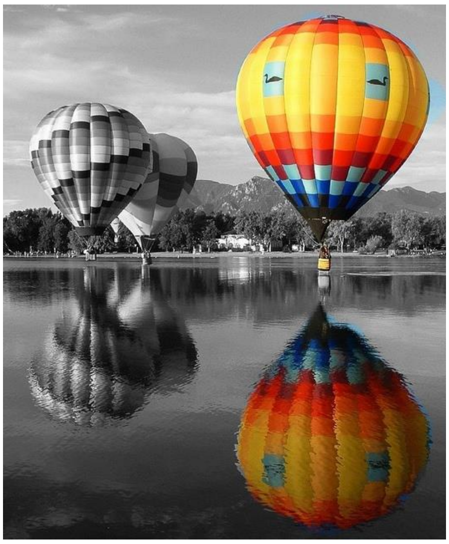


Figure 1 - Perceptual salience.

The picture represents an example of perceptual salience in a visual scene. The colored hot air balloon stands out in the grey environment and immediately attracts the attention of the observer, in a bottom-up fashion.

there is nothing special about word learning, described as a “dumb attentional mechanism” which is not even necessarily voluntary (L. B. Smith, Jones, & Landau, 1996). For example, in a study by Samuelson and Smith (Samuelson & Smith, 1998), two-year-old children played together on the floor with an experimenter, using three toys. After this, they were invited by the experimenter to move to a table to play with a fourth toy. When moved again to the floor, and asked several times to look at the “gazzer” among the toys, children selected the toy they played with on the table. According to Samuelson and Smith, the special location made the fourth toy particularly salient and thus “worth” being linked to the new verbal label.

Despite having a clear advantage of being simple and intuitive, this theory certainly has shortcomings. For example, it does not explain how children learn words that do not have a referent, such as close-class words. There is, however, another and more crucial aspect not explained by this theory: Why is one object more salient than another? Is it just for its particular properties, or is it somehow made special? Diesendruck and colleagues (Diesendruck, Markson, Akhtar, & Reudor, 2004) suggest that this is indeed the case. Using a similar experimental design as employed by Samuelson and Smith (Samuelson & Smith, 1998), the authors demonstrated that children successfully associated the verbal label “gazzer” to the toy they played with on the table only when the change in context (that is, the toy’s spatial location on the table instead of on the floor) was intentional, but not when it was accidental (that is, when the experimenter moved it to the table, but not when he dropped it on the table by accident). Further, the word was not learned when a second experimenter (who did not play with the child) named the object. The evidence collected by these authors suggest that the object was not special because of its spatial location, but because the experimenter made it special; in other words, what was important for the child was the intention communicated by the experimenter. This interpretation suggests that the attention that a child directs towards an object does not depend uniquely on the absolute saliency of the object, but more on the value that the child gives to it (L. Bloom, 1995, 2000).

1.2.2.2 The constraints theory

The constraints theory (Markman, 1990, 1991) is particularly devoted to explaining how children are able to solve the indeterminacy problem. The tenet of this theory is that children have some a-priori, innate biases concerning how verbal labels apply to their referents. The innate nature of the bias, strongly influenced by Chomsky's position on the biological pre-determination of language abilities (Chomsky, 1976), represents a crucial distinction from the previous theory; indeed, the constraints theory does not state that the biases are due to some elements being particularly salient, but instead implies that children naturally assume some kind of relations between words and the outside world (Tomasello, 2000). The whole-object bias is a paradigmatic example used by constraints theorists: When hearing "gavagai" referred to a rabbit (Quine, 1960), the child immediately assumes that the word means "rabbit" (Markman, 1990, 1991). In this theory, the role played by the caregiver is conceived as yet another external constraint: The co-occurrence of a word and its referent is considered meaningful if the caregiver transmits clear-cut signs of intending to talk about the object, thus limiting the number of possible referents (Baldwin et al., 1996). In recent years, an attempt has been made to better investigate how the constraints actually work; for example, a recent study explored the role of contextual variables, such as novelty, on the mutual exclusivity bias. As described in paragraph 1.2.1.2, this bias states that an object cannot be assigned with two different names; hence, new verbal labels are preferably assigned to new objects. However, how does a learner decide the correct referent between two objects, which do not have a name? In a preferential looking study, Mather and Plunkett presented 22-month-olds with nameless objects, which were either novel or not novel to the children. When a new word was uttered, the authors observed an increase in the children’s attention toward the novel object, as indicated by an increase in the time the children spent looking at the object compared to the familiar one. Since neither of the objects had a name, according to the mutual exclusivity bias they both represented suitable referents for the novel word. However, children decided to consistently assign the verbal label to the object they had not been familiarized with; hence, the authors concluded that novelty plays a pivotal role in the mutual exclusivity bias (Mather & Plunkett, 2012).

While the latter result does not provide any information concerning learning, it surely reflects an attempt to disengage the image of the constraints from their classical representation of something fixed and immutable. Indeed, one of the most criticized aspects of this theory is the fact that constraints are supposed to be innate and, hence, not dependent on contextual modulation. Further, this theory suffers from the same shortcomings as the association theory: If children took these biases too seriously, they would never learn anything but nouns. Indeed, all the constraints – especially the whole-object constraint – assume that the child knows that a verbal label refers to an object. But several questions then arise: How do they know when a word refers to an action? How are they able to map a complex word with a non-specific referent (for example, “breakfast”) (Nelson, 1988)? These issues are not solved by the constraint theory.

1.2.2.3 The social-pragmatic theory

The *social-pragmatic theory* of word learning (Bruner, 1975, 1983; Tomasello, 1992, 2001) has been proposed as a critical alternative to the other theories, which are criticized for not acknowledging the role played by interaction between the child and caregiver in word learning. Indeed, despite the obvious fact that language is a form of communication *between* people, “perhaps surprisingly, [the other theories] have nothing to do with pragmatics or communication" (Tomasello, 2000). Instead, as the name suggests, the social-pragmatic theory strongly emphasizes the role of shared intentionality between child and caregiver in facilitating language acquisition (Tomasello, 1992). The tenet of this model is that – while some constraints are surely needed to solve the reference indeterminacy problem – they do not need to be language-specific. The reason for this is that the process may be adequately constrained by the interactive structure of the learning setting typical for human infants, based on joint attentional activities taking place between children and their caregivers. These activities create a sort of "common ground" where it is possible to easily understand the intentions of the other person and to tune into the adult's attention; this social-cognitive ability has the advantage of allowing the child to determine the adult referent for a new piece of language without even knowing the language (Tomasello, 1992).

This theory, as compared to the association and principle theories, offers many advantages. First, it provides an explanation as to why word learning begins when it does, around 1 year of age. Why not after, when the child is better cognitively equipped? According to Tomasello, this particular period corresponds to the development of the child's ability to share attention with the caregiver. Indeed, several studies support this claim, by showing how the growth in vocabulary closely follows the development of joint attentional skills such as gaze-following (Akhtar & Tomasello, 2000; M. Carpenter et al., 1998; Tomasello, 1995). Second, the social-pragmatic theory provides explanations concerning how verbs are learned. This process is described in a study by Akhtar and Tomasello (1996), in which a child was introduced to a novel action (pushing a toy on a swing) that was always and only performed with a specific toy (Big Bird). The experimenter then introduced a novel word to describe the action, by saying "let's meek Big Bird"; however, the swing could not be found and the action was not performed. Nevertheless, children learned the meaning of the new word even though the action was never performed after the word was introduced. This result contradicts the association theory; since the action was never performed after presentation of the verb, the child did not have anything to make the link between verb and referent. The establishment of the word-action association was instead enabled by the child's ability to actively understand the caregiver's referential intention (Akhtar & Tomasello, 1996). Further corroboration for this hypothesis comes from another study, showing that new verbs are learned only when the action they refer to is performed intentionally by the speaker, but not when it is presented as an unintentional event happening accidentally (Tomasello & Barton, 1994).

Thus, the social-pragmatic theory has some advantages when compared to other word learning theories. Nevertheless, it is not completely exempt from shortcomings. First of all, despite criticizing the other theories for their partial point of view, it suffers itself from this same weakness. The idea of innate principles or simple associations is substituted by the ability of the child to read the caregiver's intentions, but little space is reserved in this theory for other possible cues. However, an important difference between this theory and the constraint theories is that while innate principles are irrefutable a-priori assumptions, the tenets of the social-pragmatic theory (for example, common ground, sharing of intentions) are assumed to follow a developmental trajectory in harmony with the child's cognitive development. Nevertheless, other problems, such as how children learn close-class words, is only vaguely resolved by invoking some social-cognitive abilities of the child, despite the fact that it is unclear how a child can infer the meaning of words such as "at" or "to" by simply sharing attention with an adult.

1.2.2.4 The emergentist coalition model

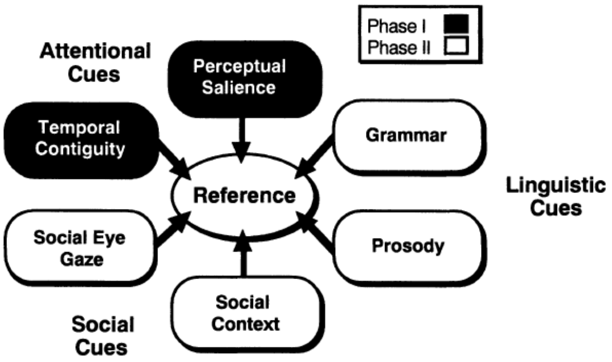
The *emergentist coalition model*, originally proposed by Hollich and colleagues (Hollich et al., 2000), represents an attempt to unify the strengths and reduce the shortcomings of the other word learning theories. Indeed, this theory arose as a critique to the partial approach expressed by associationism, constraint theory and social pragmatic theory, each focusing on a specific aspect of word learning. The *emergentist coalition model*, instead, proposes an innovative approach, different from the previous models for at least two reasons: First, it provides an integrative account of how attentional, linguistic and socio-pragmatic cues are integrated during learning; second, it emphasizes the fact that language learning is a process happening in time, not a crystalized phenomenon.

These two crucial tenets may be explained as follows: The process of word learning is divided into two phases, delineating a developmental trajectory, in which different cues (linguistic, socio-pragmatic, attentional) increase in their importance over time (L. Bloom, 2000; Hollich et al., 2000) (**Figure 2**). In the first phase, the child mostly relies on attentional cues, such as salience and temporal contiguity (see 1.2.2.1 for a more in depth explanation of these concepts), to identify possible word referents. However, as her/his cognitive and social abilities improve, the child uses more and more sophisticated cues provided by linguistic and social information (second phase). The shift from first to second phase is also reflected in the development of the link between a word and its referent, changing from being a simple association ("cow" is a sound appearing when a cow is presented) to a symbolic relation ("cow" is what the animal is). Bloom (2000) describes this developmental trajectory as "a continuum, from describing a "goes-with" relationship in an association to a decontextualized "stands-for" relationship – from an immature principle whereby a label is attached to whatever is "interesting" based on *perceptual* cues, to a mature principle when the child is apprenticed to adults who lead or direct the child's attention to a word's meaning. The continuum is from perceptually based, associationist learning to social learning" (L. Bloom, 2000). The increase in the use of social cues from the first to the second year of life, predicted by the model, may appear as somehow counter-intuitive: The more the child becomes independent, the more she/he relies on the caregiver. To explain this paradox, Bloom defines such a developmental trajectory as not specifically linked to an evolving a-priori bias in the child, but more to a complex and integrated development of social, linguistic and pragmatic skills occurring around the second year of life (Bates, 1979; L. Bloom, 1995).

The Emergentist Coalition Model comes equipped with a new method to study the specific impact of each variable it postulates. The new method, called the “Interactive Intermodal Preferential Looking Paradigm (iIPLP)”, is based on the inter-modal preferential looking paradigm (IPLP) proposed by Golinkow and Hirsch-Pacek (Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987; Hirsh-Pasek & Golinkoff, 1999). In brief, the method comprises the following: During a playing session involving two toys, a child is seated on her/his mother’s lap across the table from the experimenter. The experimenter names one of the two toys (for example, “This is a *glorp*”). Afterward, the two toys are presented together on a display board, and the hidden experimenter requests the target

Figure 2 - The emergentist coalition model.

Children shift from a reliance on attentional cues like perceptual salience (Phase 1), to a greater dependency on social and linguistic cues, like eye gaze and grammar (Phase 2; reproduced from Hollich et al., 2000).



object (Hollich et al., 2000). By changing small details, this setting allows testing for the contribution of specific social, attentional and linguistic cues; for example, to test for joint attention, the experimenter can either look at the target object or at the child or divert his gaze. The proposal of this new method represents a particularly clever way to answer a possible critique – namely that it is not trivial to study the combined effect of socio-pragmatic, attentional, and principle variables without being at risk of confounded results.

Despite the obvious advantages, a shortcoming of this model is that, once again, the theory explains very well how nouns are learned; however, entire classes of other words (prepositions and even verbs) are not considered. Recently however, the theory has been broadened to explain the principles underlying verb learning (Golinkoff & Hirsh-Pasek, 2008; Maguire, Hirsh-Pasek, & Golinkoff, 2006; Maguire, Hirsh-Pasek, Golinkoff, & Brandone, 2008). Even for this class of words, it has been hypothesized that perceptual factors are predominant in the first stages of learning, leading interesting and immediate events to be preferred as verb referents. However, as the meaning of a word may still be nebulous, social and linguistic information intervene to solve the indeterminacy and provide referents for more abstract verbs.

Lastly, the most concerning weakness in this model is that the child has a predominantly passive role. As expressed by Bloom, the child has "relatively passive roles as a *perceiver* of physical cues, the *receiver* of social cues, and somehow the *possessor* of constraints, biases, or heuristic principles that filter the available information for the child" (L. Bloom, 2000). However, anyone who has come across a child in his life may immediately recognize a passive child as an incredibly rare phenomenon.

1.2.2.5 The social gating hypothesis

In recent years, the role played by social interaction in an individual's cognitive processes has been dramatically emphasized in several fields of cognitive neuroscience, including learning. Indeed, one of the most recent proposals in this domain, the *social gating hypothesis* (Kuhl, 2007) claims that learners (both infants and adults) may be predisposed to pay particular attention to linguistic information when they are presented in a social context (Sage & Baldwin, 2010), in which the caregiver often adapts his behavior to the learner in a pedagogical fashion (Csibra & Gergely, 2009)[9].

Despite having broad applications for learning in general, this theory was originally formulated to explain early language learning and, more specifically, phonological acquisition and phonetic discrimination (Kuhl et al., 2003, 2008). But how does social gating work? According to Patricia Kuhl (Kuhl, 2007; Sage & Baldwin, 2010), a social agent conveys two characteristics making her/him particularly salient – she/he *interacts* with the child, and does so in a *contingent* manner; in other words, she/he adapts to the child with a punctual and responsive attitude (K. Bloom & Esposito, 1975; K. Bloom, Russell, & Wassenberg, 1987). Thus, it is not only the presence of another person that simply heightens the attentional threshold of the learner, but the fact that the other person is directly involved in the learning process (for a more in depth discussion on the distinction between actor and observer in social contexts, see the Methods section). In support of this hypothesis, Bloom and colleagues (K. Bloom et al., 1987) found that children's tendency to emit vocalizations is enhanced when they are engaged in interactive turn-taking with a responsive caregiver, as opposed to a randomly responsive adult. This evidence was more recently confirmed for word learning in a study by Pereira and colleagues, showing how interactive turn-taking in toddler-parent dyads was positively correlated with an increase in the child's vocabulary (Pereira et al., 2008).

Despite being born in the domain of phonological learning, the social gating hypothesis has potential applications even in the domain of word learning. In

---

[9] According to Sage and Baldwin (2010), pedagogy might be thought of as a particular form of social gating, in which particularly relevant is the reciprocal understanding and decoding of the other's intention.

particular, it has been shown how the co-occurrence of a word and its referent is considered meaningful if the caregiver transmits clear-cut signs of intending to talk about the object (Baldwin et al., 1996). In the study by Baldwin and colleagues, infants were confronted with two similar word-learning situations, in which a new word was assigned to a toy. Crucially, only in one condition could the infant infer social cues (for example, gaze direction) from the experimenter, as in the second condition the experimenter was behind a piece of rice paper[10]; children only associated the name with the toy in the first condition. In a subsequent study, the authors also discarded the hypothesis that learning is improved by social cues simply because they make the situation more salient. The study investigated causal learning in a non-linguistic domain (learning how to use a tool) involving a social and a non-social condition. The authors observed that infants were equally attentive in both conditions; nevertheless, only infants in the social condition successfully learned the correct usage of the tool (Sage & Baldwin, 2011). The latter experiment not only rules out the hypothesis that general arousing mechanisms underlie socially gated learning, but provides evidence that this type of learning applies to domains other than language (see Moses, Baldwin, Rosicky, & Tidball, 2001, for an example on the emotional response to an object).

The social gating hypothesis has several points in common with the social-pragmatic theory of word learning. However, two crucial differences can be identified. First, the social-pragmatic theory of word learning is prevalently language-based and language-directed, while the social gating hypothesis is potentially domain-general; second, the social pragmatic theory postulates language learning to be dependent on high-level cognitive processes (such as mentalizing; Frith & Frith, 2006), while the social gating hypothesis offers the advantage of explaining a complex phenomenon by employing relatively

---

[10] A possible criticism might be that the experimenter behind the rice paper represents an unusual situation, requiring more attention. However, the authors assessed visual attention to the target toy and found it equivalent in both conditions.

simpler cognitive processes, such as attention, and properties of the stimulus, such as saliency. In this way, the social gating hypothesis brings language learning back to the more general category of communication, allowing a comparison with other species in which communication is socially gated, such as birds (Heyes, 1994; Janik & Slater, 2000; Laland, 2004). In contrast, the social-pragmatic theory states that even our closest relatives fail in social cognitive tasks, and this is why they do not develop language (Tomasello & Carpenter, 2007).

To some extent, the social gating hypothesis may thus be seen as an attempt to unify association theories with social-pragmatic theories. Despite being originally formulated for the early stages of word learning (that is, phonetic learning), its potential as a general mechanism for learning has also been outlined (Sage & Baldwin, 2010).

1.2.2.6 Remarks on word learning models

In the previous paragraphs, the most influential models of word learning have been described, together with a more general model describing language learning as a socially mediated mechanism. Each of the theories reviewed above have both strengths and weaknesses; however, some general remarks can be put forward.

The first three models (the association theory, the constraint theory and the social pragmatic theory) have long competed with each other for the title of “true” word learning theory. All of them are based on a specific partition of the word learning process, and leave little to no space for the relevance of information other than the one being focused on. In particular, the constraint theory appears particularly problematic, as it postulates innate language-related principles that can be applied only to language; in this sense, it is purely a linguistic theory without psychological implications (Nelson, 1988). The reflection upon the partiality of these theories led to the development of the integrative view proposed by the emergentist coalition model. Not only does this model have the advantage of proposing a more comprehensive theory, but it also stimulates reflection upon the importance of defining an appropriate method to test a specific model and hypothesis. If a model is new, it requires a new method, created ad hoc to test the novel hypotheses it generates. The studies described in this dissertation focus on exploring the importance of social interaction to model the way new words are learned, which corresponds to a novel research question (see chapter 2). Accordingly, a new method specifically tailored to test the hypotheses generated by this research question has been developed and it is described in chapter 3. Two validation studies are described in the "Studies" section.

Lastly, it must be emphasized that all of the theories have points of overlap. An important example is that all social theories emphasize the fact that for the child, it is not sufficient to just have a person being present; what is needed is a responsive, interactive caregiver. The punctuality of the interaction, reflected by turn-taking rituals, represents a basic prerequisite for developing common ground, in which learning is possible. Turn-taking is a process by which people in an interactive situation decide whose turn is next, with a general rule being that both overlaps and silences should be minimized (Stivers et al., 2009). Turn-taking rules not only represent a universal of human interactions, but they also develop from a very early age (K. Bloom et al., 1987; M. Wilson & Wilson, 2005). The combination of these two aspects (universality and early development) suggests that turn-taking may support the exchange of information during communication that is necessary for learning. This claim is supported by evidence that turn-taking is involved in several aspects of language acquisition, from very basic vocalizations (K. Bloom et al., 1987) to word learning (Pereira et al., 2008). But how do people know how to "take turns"? M. Wilson and Wilson (2005) proposed that the speaker's rate of syllable production determines endogenous oscillators in the brains of the speaker and the listeners become mutually entrained (that is, coupled), but counter-phased. In brief, this means that the potential for the listener to start talking is maximal when the talker phase of the cycle is minimal. Crucially, Wilson and Wilson suggest that other activities that are complex, rhythmic, and interpersonally coordinated may be governed by similar principles. A prototypical example of these activities is represented by music (Jungers, Palmer, & Speer, 2002; Large & Jones, 1999; Large & Palmer, 2002), but other types of socially interactive situations, such as sitting side by side in a rocking chair, also elicit synchronous oscillatory sways (Richardson, Marsh, Isenhower, Goodman, & Schmidt, 2007). While this interpretation of turn taking is particularly powerful in bringing together pieces of evidence from different fields, it also has the advantage of explaining why temporal coordination expressed by turn-taking should facilitate learning: As the endogenous oscillators become coupled, modification in the timing of one of the two would lead to changes in the others. According to Kuhl's proposal (Kuhl, 2007), this is one of the reasons why a social partner is particularly salient for the child: "he *interacts* with the child, and he does so in a *contingent* manner" (Kuhl, 2007; Sage & Baldwin, 2010). In addition to the psychological implications of this evidence (for example, the child may be more motivated to learn when the caregiver is responsive), this has direct implications for attention. Indeed, as association theories suggest (Balsam et al., 2010; L. B. Smith, 2000), the time of presentation between word and referent is critical for ensuring the establishment of a link between them. The coupling between the endogenous oscillators of child and caregiver facilitate this by creating a common ground that is not only psychological in nature, but also based on precise timing. In turn, this "common psychological and temporal ground" would ensure that when the caregiver utters a new word the child will be maximally predisposed to receive it, thus allowing a prompt mapping with the referent.

It is crucial to note that this interpretation is not limited to children; indeed, evidence of temporal coordination occurring during social interaction in adults is abundant (Demos, Chaffin, Begosh, Daniels, & Marsh, 2012; Kawasaki, Yamada, Ushiku, Miyauchi, & Yamaguchi, 2013; Oullier, de Guzman, Jantzen, Lagarde, & Kelso, 2008; Richardson et al., 2007). However, as the role of social interaction in adult word learning has not been investigated as of yet, the possibility that temporal coordination dynamics may be involved remains to be explored.

1.2.3 Is second language learning different from first language learning?

> *"The Babel fish is small, yellow, leech-like, and probably the oddest thing in the universe [...] If you stick one in your ear, you can instantly understand anything said to you in any form of language: the speech you hear decodes the brain wave matrix."*
>
> Douglas Adams

The existence of a universal translation device such as the Babel fish would make everyone's life a lot easier, in a world becoming increasingly interconnected and multi-cultural. Unfortunately, such a "device" has not yet been invented, leaving only one option to those who want to understand a foreign language: To learn it.

In the previous paragraphs, the processes underlying language learning have been discussed with a particular focus on the acquisition of a first language. To accomplish this task, the learner has to segment the continuous flow of sounds (speech) into separated and specific units (words); subsequently, she/he has to identify what the words are referring to and to learn which referents may be grouped under the same verbal label (referencing and generalization problems).

While children are surely the most remarkable learners, the ability to learn a new language does not end in childhood. As pointed out in the Babel fish example, learning (at least) a second language is fairly common in the modern world. But how is learning a second language different from learning a first language? Are they the same process? Or are they completely different? These questions reflect the doubts faced by the first researchers of bilingualism. Before the beginning of the imaging era, psychological and clinical data suggested language learning to be a process limited to infancy and childhood; in other words, it was hypothesized that learning a second language in adulthood was a different (and incredibly less efficient) process as compared to learning a first language as an infant. Imaging methods significantly contributed to changing this hypothesis by showing evidence of remarkable similarities in language learning in children and adults. Further, important modulating variables were identified (age of acquisition, exposure, proficiency), leading to models of the bilingual lexicon significantly improving over the years. The most relevant historical and theoretical progresses in the study of bilingualism are described more in depth in the following paragraphs.

1.2.3.1 The sooner the better: The Critical Period Hypothesis

Even to the present day, it is fairly common to hear the claim that the older one gets, the more difficult it becomes to learn a foreign language. This common saying is supported by a long-standing hypothesis, stating that only when a language is learned during early childhood it is possible to attain native-like proficiency. This limitation supposedly depends upon the existence of a specific time window during which language learning is possible. This Critical Period Hypothesis was formulated around the middle of the 20th century (Lenneberg, Chomsky, & Marx, 1967; Penfield & Roberts, 1959), at the peak of the debate on the nature of language (Chomsky, 1976). In this context, the Critical Period Hypothesis represented an attempt to ground language in the biology of the brain together with sensory modalities, for which the existence of sensitive periods has been extensively demonstrated (for example, vision; Hensch, 2005). More specifically, the Critical Period Hypothesis for language identifies a specific time window, opening at around 2 years of age and closing with puberty, during which the mere exposure to speech is sufficient to acquire the mother tongue. However, if the critical period closes without the child having had adequate input, then he will no longer be able to master a language in a native-like manner (Krashen, 1973; Lenneberg et al., 1967). The upper boundary of the critical period is defined by the neurobiological and neuropsychological constraint imposed by the lateralization of language in the left hemisphere, which is supposed to be completed in puberty (Krashen, 1973; Lenneberg et al., 1967). Evidence from aphasic children supports this claim: Indeed, right hemispheric lesions cause aphasia more frequently in children than in adults; further, aphasic children usually recover more easily than adults after left hemispheric lesions (Basser, 1962).

Interestingly, one of the most cited and discussed pieces of evidence in support of the Critical Period Hypothesis, the case of feral children, is the one supporting it the least. “Feral (or deprived) children” is a label used to indicate cases of infants deprived of any social and linguistic input from birth or early infancy onwards. Usually, these children do not develop a functional language and, in the worst cases, they do not acquire language at all. The most studied case is the one of Genie (Curtiss, 1977), a girl rescued at age 13 after surviving 11 and a half years of almost complete social and sensory deprivation. During this time, nobody ever talked to Genie and any attempt she made to vocalize was severely punished. Her complete lack of linguistic abilities at her discovery, together with the fact that she never really acquired a fully functional language even after years of teaching, supports the Critical Period Hypothesis.

However, the environment Genie grew up in was deprived of everything, not only language; as such, it is impossible to explain her deficits purely on the basis of a lack of linguistic exposure. Further, it must be noted that Genie gained at some stage some basic linguistic abilities; more specifically, she was able to learn new words, despite a deficient mastering of morphology and syntax (Fromkin, Krashen, Curtiss, Rigler, & Rigler, 1974; Krashen, 1973).

More convincing evidence supporting the Critical Period Hypothesis comes from deaf children and adults acquiring American Sign Language (ASL) (Johnson & Newport, 1991; Newport, 1988, 1990). ASL is particularly well suited to study of the critical period hypothesis, since only a small proportion of deaf signers (5/10%) learn a language from birth; the remaining percentage of deaf signers learn later on in life, often well after childhood (Newport, 1990). Given this range, ASL provides a unique case to clearly evaluate the effect of age of acquisition (AoA) on proficiency in the first language. Newport and colleagues studied three cohorts of deaf signers with different age of acquisition of ASL as their primary language: Native learners (exposed from birth), early learners (exposed from 4-6 years of age) and late learners (exposed after 12 years of age). Supporting the hypothesis of a critical period, Newport and Supalla (1990) observed a clear effect of age of acquisition on ASL proficiency, as "the later the language is learned, the less its use is native in character" (Newport, 1990). Despite their apparent robustness, these results have nonetheless been challenged on the ground that they are likely confounded. Indeed, the transition between childhood and puberty is characterized by major changes in the learning environment, which are very difficult to control for. As an example, young children are immersed in a social environment shaped toward the learner and her/his necessities, while older learners are most often not nearly supported as much (Bialystok, Craik, & Luk, 2012; Bialystok & Hakuta, 1999).

The Critical Period Hypothesis has been hypothesized to affect only first language learning, but also to extend to second language learning (Long, 1990). To test this hypothesis, Johnson and Newport evaluated syntactic and morphological competence in adult learners of English as a second language for native speakers of Korean or Chinese (Johnson & Newport, 1989; Newport, 1990). Again, these authors found a strong relationship between the starting age of learning and performance in a test of linguistic competence. More specifically, test performance was linearly related to starting age until puberty; after puberty, performance was in general lower, but highly variable and independent from starting age.

Taken together, this evidence seems to support to some extent the existence of a critical period. However, while AoA effects on language performance have been consistently replicated, the fact that they reflect the existence of a critical period has been challenged (Birdsong, 2014; Birdsong & Molis, 2001). Indeed, the hypothesis that a sensitive time window exists for language acquisition would predict a) a sudden drop in language learning abilities after puberty, b) the impossibility to find late learners with native-like performance, and c) an independence of age-related effects from the typological relation between L1 and L2 (Birdsong & Molis, 2001). These three hypotheses have been the focus of extensive research, which has not necessarily provided evidence in favor of the Critical Period Hypothesis. Indeed, with regard to the first hypothesis, several studies showed that L2 performance clearly depends on L2 AoA, with a significant negative correlation between AoA and performance (Birdsong, 1999; Flege, Yeni-Komshian, & Liu, 1999; Piske, MacKay, & Flege, 2001); however, there is no marked discontinuity in the pre-post puberty period. Instead, AoA effects are evident even when the L2 is learned after puberty (Bialystok & Hakuta, 1999). For example, Hakuta and colleagues demonstrated this by analyzing a corpus of data from 2.3 million Spanish or Chinese immigrants learning English (Hakuta, Bialystok, & Wiley, 2003). As for the second hypothesis, some studies demonstrated the possibility of attaining native-like levels of proficiency even in later learners. Bongaerts, for example, reported the case of Dutch late learners of English or French who were judged to have a native-like pronunciation by a native speaker of the L2 (Bongaerts, 1999; Bongaerts, Mennen, & Slik, 2000). Lastly, the Critical Period Hypothesis predicts similar effects independently of the typological distance between L1 and L2. However, studies conducted on learners with different L1-L2 pairs tend to refute this hypothesis; for example, the results from Johnson and Newport (1989) with Korean/Chinese speakers learning English were disconfirmed in the study by Birdsong & Molis, in which Spanish speakers learned English (Birdsong & Molis, 2001; Johnson & Newport, 1989). This contrasting evidence suggests that the distance between L1 and L2 may play a role in attaining L2 proficiency that is not predicted by the Critical Period Hypothesis.

Taken together, these results suggest that while first and second language learning abilities do decline with age, there is no fixed time window in which learning a language is possible. Rather, a complex interplay of cognitive and social factors (including contextual factors, linguistic distance between L1 and L2, AoA) should be held responsible for the observed patterns of attained proficiency.

1.2.3.2 How to fit two languages into one brain: neural organization in bilinguals

A last prediction that may still "save" the Critical Period Hypothesis concerns the way L1 and L2 are organized at the neural level. Indeed, if the Critical Period Hypothesis is correct, then the neural organization subtending the usage of the second language should be different from the first language, as different mechanisms are employed to learn them. Instead, if the learning mechanism is the same for both languages, then their neural representations should be fairly similar.

Until the beginning of the 21st century, the investigation of how the brain handles multiple languages relied mostly on clinical case studies of bilingual (or polyglot) aphasic patients (Fabbro, 2001). These studies provided an heterogeneous picture: Frequently, a brain lesion impairs one language but not the other, but sometimes both languages are affected to some extent; at times, the lesion impacts prevalently L1, other times L2. It was not until the '70s that these puzzling results started to be explained, thanks to the first cortical stimulation studies (Ojemann & Whitaker, 1978), and these investigations began to unveil the principles underlying the neural organization of the bilingual brain. The language system of bilingual patients includes some areas in common between L1 and L2, while other areas are specific for one or the other language. Further, L2 has – as a rule of thumb – larger representations than L1 (Abutalebi, Cappa, & Perani, 2001). These first observations in clinical populations have been subsequently enriched by several years of neuroimaging research in healthy participants. These studies revealed that, indeed, from the point of view of language production, there are no differences in the pattern of activation observed for L1 and L2. However, this is true only for early bilinguals and late bilinguals with high L2 proficiency. In the case of late bilinguals with low levels of L2 proficiency, a broader network is recruited for L2 as compared to L1, with high variability in activations observed in different participants. During comprehension tasks, the pattern of neural activation for L1 and L2 is remarkably similar in highly proficient speakers regardless of their AoA; instead, low proficient speakers display significantly more extended activations when listening to L1 as compared to L2[11]. The common network includes classical left-sided perysilvian areas including superior and middle temporal gyri, the angular gyrus, temporal pole, and the middle and inferior

---

[11] The different extension of the network recruited by L2 during comprehension (L2 less widespread than L1) and production tasks (L2 more widespread than L1) in low-proficiency bilinguals has been hypothesized to reflect the recruitment of additional resources in the first case (to produce an output) and a more limited elaboration of the input in the latter (Abutalebi et al., 2001).

frontal gyrus, while no clear pattern is identified for low-proficient participants (Abutalebi et al., 2001). These results are independent of the linguistic distance between L1 and L2; in other words, highly proficient bilinguals show very similar patterns of activation regardless of the similarity between the two languages, as demonstrated in studies in which L1 and L2 were, respectively, Italian and English or Catalan and Spanish (Perani et al., 1998) but also more distant languages such as English and Mandarin (Chee et al., 1999).

Data on the neural correlates of L1 and L2, thus tend to disconfirm the Critical Period Hypothesis. However, the evidence reported above confirms that age of acquisition indeed represents a pivotal factor in word learning, together with proficiency. In recent years, another variable has been added to the pool of modulatory factors – exposure. Perani and colleagues provided the first evidence for the role of exposure in a group of highly proficient and early-acquisition speakers of Spanish and Catalan. All participants were exposed daily to both Spanish and Catalan in social contexts; however, they differed in the amount of daily L1/L2 exposure. This selected group of bilinguals was monitored with fMRI while performing a phonemic verbal fluency task. The study yielded two important results: First, age of acquisition was confirmed to be a crucial factor, even for very early learners, who acquired L2 before the age of 3. Indeed, the production of words in L2 recruited a more extended network as compared to language production in L1. Second, the effect of AoA was modulated by exposure: L2 activations tended to be more similar to L1 activations in bilinguals who are more exposed to L2 during their daily life (Perani et al., 2003). More recently, these authors reinforced this claim by showing that when proficiency and exposure are kept high, the neural network supporting L2 is the same as for L1, independently from AoA (Consonni et al., 2013).

Electrophysiological investigations have provided consistent evidence: In general, the later a second language is acquired the lower the level of attained proficiency (Mueller, 2005; Weber-Fox & Neville, 1996). As for the processing of semantic information, the N400 effect (see paragraph 1.4.1 for details) has been found to be remarkably similar in shape across first and second languages, but often delayed in bilinguals when processing L2 depending on AoA and exposure (Ardal, Donald, Meuter, Muldrew, & Luce, 1990; Kotz & Elston-Güttler, 2004; Moreno & Kutas, 2005).

Taken together, the evidence reviewed above points toward strong similarities in the network subtending first and second languages, in both production and perception. The differences between the two languages seem to be related not to the usage of different learning mechanisms, but more to the relevance of factors such as the age of acquisition, exposure and proficiency. Despite the similarities, a difference seems to emerge between L1 and L2: While the semantic system seems to be less influenced by AoA, morphosyntactic processing appears to be more affected. An explanation for this effect may be that, while it is reasonable to assume that the first is mediated by similar mechanisms in L1 and L2, the second is likely mediated by other mechanisms involved in more procedural learning (Abutalebi, 2008). The next paragraph discusses the details of this assumption.

1.2.3.3 The Declarative/Procedural and Convergence Hypotheses for L2

While the Critical Period Hypothesis has been significantly reshaped by neuroimaging evidence, the debate concerning which mechanisms are employed in different aspects of L2 learning is still ongoing.

A first hypothesis focuses on clarifying why morphosyntax aspects seem to be more influenced by AoA than lexical aspects. A possible explanation for this evidence relies on the way they are acquired: While first language is learned spontaneously, second language is in most cases taught explicitly in scholarly settings (Paradis, 1994, 1997, 2004). The consequence of this difference in the learning procedure is that grammatical and syntactical rules are learned via different circuits in the adult – who have explicit knowledge of them – compared to children, who know them only in practice but do not have conscious access to them. This hypothesis, known as the *Declarative/Procedural Hypothesis* (Ullman, 2001, 2005), predicts that L1 (mostly in the morphosyntactic aspect but to some extent also for the lexicon) may hinge more on a *procedural* system – relying on fronto-striatal networks, while L2 might be more dependent on an explicit *declarative* memory system hinging on left temporal regions (Ullman, 2001). The evidence reviewed in the previous paragraph, however, tends to demonstrate that L1 and L2 are mediated by highly overlapping substrates, thus weakening this hypothesis.

An alternative explanation is that L2 acquisition is based on an already specified L1 network, and receives convergent neural representation within the representations of the language learned as the L1 (Green, 2003). In other words, this hypothesis, known as the *Convergence Hypothesis,* states that “the acquisition of a second language will utilize existing devices and the particular properties of the language will affect the relative demands on these devices” (Green, Crinion, & Price, 2006). The Convergence Hypothesis also explains the disappearance of the differences between L1 and L2 as a function of proficiency, as the neural representation of L2 will ultimately converge to L1 as the proficiency increases (Abutalebi, 2008; Green et al., 2006). Importantly, this hypothesis states that also grammar and syntax might be mediated by the same mechanisms supporting L1 learning. Support for this theory come from studies on artificial grammar learning: For example, Opitz and Friederici (2004) have successfully shown how an increase in proficiency in the rules of an artificial language positively correlates with activation of Broca’s area, a region involved in syntactic processing in L1 (Opitz & Friederici, 2004). The convergence hypothesis has the advantage of explaining another important observation about L2 neural organization; as previously described, the network supporting L2 production for low proficient bilinguals is, in general, broader than for L1. The extra areas recruited in L2 sometimes include more frontal areas than those usually involved in language, such as left prefrontal cortex, ACC, and basal ganglia, involved in cognitive control (Abutalebi, 2008; Miller & Cohen, 2001). Cognitive control is an essential ability for bilinguals, who face the problem of competing L1 and L2 names for the same objects. As it will be explained in the next paragraph, most of the theories on the distribution of the lexicon in bilinguals assume that two lexical representations (one for each language) access a common semantic storage (see paragraph 1.2.3.4). Hence, the problem arises: What happens when a concept is activated? How is the access to the two lexical representations controlled? It has been suggested that this competition can be resolved by *inhibiting* the language not required at the moment (Green, 1998; Rodriguez-Fornells, De Diego Balaguer, & Münte, 2006). Low-proficiency bilinguals, compared to high-proficiency bilinguals, will have more difficulties in mastering this switch between languages, calling into play the areas involved in cognitive control (Abutalebi, 2008; Abutalebi & Green, 2008; Rodriguez-Fornells et al., 2006).

1.2.3.4 Models of the bilingual lexicon

The experimental evidence from bilingualism studies points towards the existence of a hierarchical organization in the bilingual lexicon, with a single conceptual storage for both L1 and L2 words but separate lexical levels (Green, 2003; Green et al., 2006). The question remains open, however, as to how these representations are connected. While agreeing on the idea that representations may be hierarchically organized, theories developed to model the distribution of the lexicon in bilinguals differ mostly on how these representations are connected.

The first two theories, the Word Association (WA) and Concept Mediation (CM) theories (Potter, So, Eckardt, & Feldman, 1984), have been developed to test two theoretical alternatives: Either the two languages are connected at the lexical level (WA) or through a common conceptual representation (CM) (Dufour & Kroll, 1995). The WA model postulates that L1 and L2 words are directly linked (or associated) to each other, but only L1 has access to the semantic storage; instead, the CM model holds that the two lexical representations are directly connected to the semantic storage, but not connected with each other. Potter and colleagues (Potter et al., 1984) investigated these two alternatives by testing bilinguals in a picture naming and translation test. The picture-naming task was used since, in order to name a picture, it is first necessary to access the semantic storage to extract the concept depicted in the image. Thus, if both picture naming and translation access the semantic level, then the two tasks should elicit very similar response time patterns. Instead, a difference between the two would speak in favor of the WA model, which does not require access to semantic representation. Potter and colleagues observed very similar response times for picture naming and word translation; hence, they conclude that Concept Mediation was a more accurate model of the bilingual lexicon.

Despite its intuitive appeal and the evidence provided by Potter and colleagues, the model was quickly criticized on the grounds that it could not account for effects of proficiency. Indeed, it was shown that while highly proficient bilinguals are able to access concepts directly from L2 as predicted by the CM model, less proficient bilinguals rely on direct associations between L1 and L2 words, as predicted by the WA model. This conclusion was based on evidence that, for low-proficiency bilinguals, picture naming in L2 was slower than translation from L1 to L2 (Kroll & Curley, 1988). How can one reconcile these results? A possible solution to this problem was to combine the CM and WA models in a more comprehensive, unified model. The Revised Hierarchical Model (RHM, **Figure 3**) (Kroll & Stewart, 1994) was developed under this premise as a model of word production able to explain the translation performances of late (less proficient) bilinguals. The explanation that the RHM provides for this phenomenon is that the links between words and concepts in L1 and L2 have different strengths: While L1 has privileged access to meanings via strong connections, L2 only has weak links with the semantic storage and thus requires mediation via the corresponding translation in L1. Thus, while the L2 → L1 translation may be accomplished without semantic access, the L1 → L2 translation always requires semantic access (Kroll, van Hell, Tokowicz, & Green, 2010). Empirical evidence supports this theory by showing how forward translation (L1 to L2) was more likely to recruit the semantic system compared to backward (L2 to L1) translation (Kroll & Stewart, 1994; Sholl, Sankaranarayanan, & Kroll, 1995). In its revised form, the RHM model assumes that the weak link between L2 and lexical representations is asymmetrical and stronger for recognition compared to production. This change in the model was based on evidence showing that production in L2 changes most dramatically with proficiency (Kroll et al., 2010).

The RHM model has recently been challenged on the basis of five distinct arguments (Brysbaert & Duyck, 2010), to which Kroll and colleagues promptly responded (Kroll et al., 2010). First, there is little evidence for separate lexicons; second, there is little evidence for language selective access; third, including excitatory connections between equivalent lexical translation risks

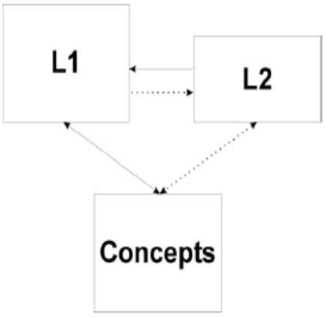


Figure 3 - Revised Hierarchical Model (RHM).
The model postulates a single semantic storage and two distinct lexical representations, one for the mother tongue (L1) and one for the second language (L2). L1 and L2 are interconnected and connected with the semantic storage. However, L2 has only weak connections with it; L1 has direct, strong connections with the semantic storage and weak connections with the L2 lexicon. Reproduced from Brysbaert & Duyck, 2010.

impeding word recognition; fourth, the connections between L2 words and their meanings are stronger than proposed in RHM; fifth, it may be necessary to make a distinction between language-dependent and language-independent semantic features. Regarding the latter point, it must be pointed out that despite two-level storage being the most common hypothesis, some authors suggest that three levels should be postulated instead; Paradis (Paradis, 1997), for instance, proposes a three-levels approach with separate lexical, semantic and conceptual storage. In his view, only the conceptual storage is shared between languages, while lexical and semantic accesses are language-specific. To overcome the limitations observed in the RHM, Brysbaert and Duyck propose to modify the model towards a *Distributed Feature Model* (DFM), in which proper weights are assigned to the connections between lexical and conceptual levels, so that L2 words activate fewer semantic nodes than L1 words (Brysbaert & Duyck, 2010).

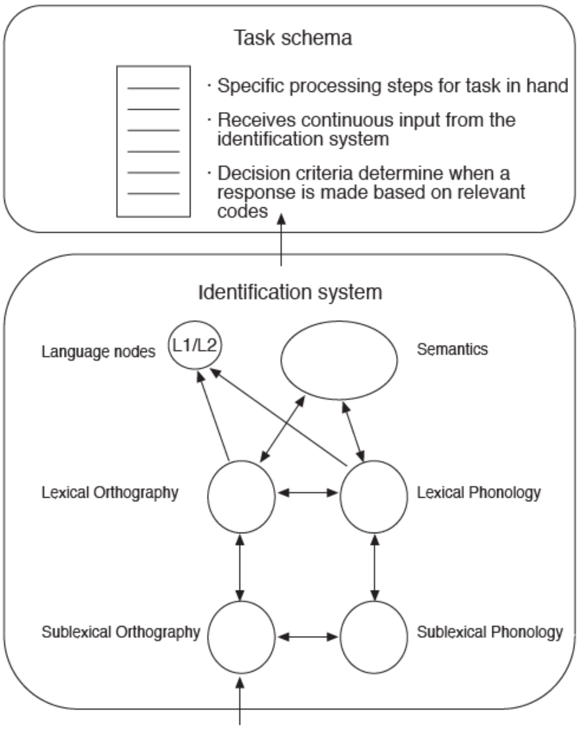
Lastly, the *Bilingual Interactive Activation* (BIA) model of visual word processing is an entirely new model of the bilingual lexicon, developed by Dijkstra and colleagues (Dijkstra & Van Heuven, 1998; Dijkstra, Van Heuven, & Grainger, 1998; van Heuven, Dijkstra, & Grainger, 1998). Based on an entirely new rationale, the BIA was created by extending a previously available model of the monolingual lexicon, namely the Interactive Activation (IA) model (McClelland & Rumelhart, 1981). This approach is novel when compared to the CM or WA that were developed to explain existing data. Further, the BIA is much more detailed in comparison to its predecessors, as it contains not only lexical and semantic information but also orthographical and, in its latest revision, phonological information (BIA+, **Figure 4**) (Dijkstra & Van Heuven, 2002).

This model postulates the existence of two different mechanisms, an identification system and a talk schema. The identification system is used to decode words: Phonological and orthographical information corresponding to the word form is originally activated by the input and, in turn, accesses the semantic and lexical levels. The talk schema represents the real novelty of this approach, as they imply that the information identified by the identification system is influenced by the context of usage, including, for example, instructions or task demands.

To summarize, models of the bilingual lexicon have become progressively more complex and comprehensive in a similar way to the models of first

Figure 4 - The upgraded Bilingual Interactive Activation (BIA+) model.

The model includes ortographic, phonological and semantic information. Further, two systems are postulates: the first system (bottom box) is the word identification system used to decode words; the second system (upper box) is a decisional system in which information is used in relation with the contextual task demands (reproduced from Dijkstra and Heuven, 2002).



language learning. In particular, contextual information has acquired the status of a specific factor influencing the usage of the language, together with traditional neuropsychological variables.

1.2.3.5 Remarks on Second Language Learning

The previous paragraphs provided a description of how second language is learned and mastered by early and late learners, with evidence from neuropsychology and neuroimaging. Years of research, started around the middle of the last century, have significantly advanced our knowledge of how a second language is learned. The studies conducted so far provide consistent evidence that second language learning is not particularly different from L1 learning. Evidence of this claim can be summarized as follows: First, there is no critical period for language learning; a second language can be learned at any point in life. Nevertheless, the mastering of a second language is influenced by various factors, most notably age of acquisition, proficiency, and exposure. Second, L2 is acquired with the same neural learning mechanisms as L1, but – once again – AoA, proficiency, and exposure modulate its usage and lead to the involvement of additional brain regions, in particular control areas. Third, models of the bilingual lexicon have progressively become more comprehensive, pointing towards a global description of how second language is used including the context of use.

One important piece of evidence needs to be highlighted: Despite never being clearly named as "social information", the role played by social context on second language learning has been clearly outlined throughout this chapter. For example, it has been demonstrated how exposure to a language dramatically influences the speaker's proficiency, even in late learners. This exposure was identified in the reported study as the amount of time learners were presented with L1 or L2 at home or outside, during social situations (for example, Perani et al., 2003). This description suggests that the term “exposure” entails a complex set of variables, in which the amount of input cannot be the only relevant aspect. Nevertheless, as suggested by Tomasello for first language learning, the input must come from someone (Tomasello, 2000). For adult learners this is not entirely true; in this case, the input may also come from somewhere else (for example, from a book the learner is reading). While adults – in comparison to children - can certainly have speech inputs from sources other than fellow humans, social interaction still represents a primary venue for learning that needs to be investigated. Indeed, as stated in the previous paragraphs (see, in particular, paragraph 1.1 and paragraph 1.2.2) language is primarily a form of communication employed to convey meanings between people. An interest in this line of enquiry is further justified by the extension of the latest model on the bilingual lexicon to include the context of use, which refers to the specifics of the task at hand. The addition of this module testifies a shift of focus from a perspective purely based on the learner, to a model which takes into account the relation between the learner and its environment, including other people. A further indirect acknowledgment of the possible role of social interaction on second language learning comes from the case of Genie. Although discussed in relation to the Critical Period Hypothesis, the data on Genie’s linguistic abilities can be observed from another point of view; during her captivity the young girl was not only deprived of language, but of any social contact. Conversely, her improvements in language happened in a social environment, where she bonded with researchers and psychologists working on her case. Lastly, Bialistock, Hakuta and colleagues (Bialystok & Hakuta, 1999; Hakuta et al., 2003) suggest that the reduced ability to learn a language may depend on the type of social support received by the learner; this explanation helps to enlighten the role of exposure on language processing, despite the fact that the type of training – implicitly by immersion or explicitly by teaching – does not elicit differences in language processing in late L2 learners (Batterink & Neville, 2013). If the advantage provided by exposure does not depend on the type of learning per se, the role of other factors (for example, social support) should be critically investigated.

Asserting that social context should be considered as a relevant factor for language learning may seem like stating the obvious. Nevertheless, research in this area is astonishingly lacking. This observation has provided the motivation for the studies described in this dissertation. In the following chapters, the aim and research questions for this research are outlined.

2

Research Questions

“*When you hear hoof beats, think horses not zebras.*”

Theodore Woodward

The aim of this dissertation is to deepen our knowledge of the mechanisms supporting word learning, in healthy adults during social interaction. As described in the previous chapter, the role of social context in language learning has been frequently hypothesized as pivotal. Studies in L1 acquisition have theoretically and empirically supported this notion, yet no systematic investigations have been conducted in adult learners. The studies presented in this thesis therefore focus on this neglected factor in second language learning. The importance of exploring this aspect arises from several lines of evidence, whose theoretical bases have been explored in the previous chapter.

The literature reviewed so far has shown how the processes underlying learning are remarkably consistent across different levels of language expertise; studies on L2 acquisition highlight that the acquisition of a new language is possible in adulthood, and that it is largely supported by the same brain mechanisms used by L1. This evidence contradicts the Critical Period Hypothesis: Indeed, although age of acquisition represents an influential factor in the attained proficiency, mastering a second language in adulthood critically depends upon the context of use, that is, the amount of exposure and usage of the new language during daily activities.

Hence, if similar processes support vocabulary learning in adults and children, the question arises as to whether they are also influenced by the same variables. The role of exposure in shaping the language abilities of L2 learners suggests that adults may not be entirely self-sufficient learners; rather, they may be more similar to children than previously thought. In this regard, a theoretical comparison with L1 models may be particularly helpful to understand how social interaction, an aspect largely neglected in adult learning, may exert its influence. Other authors have employed this type of approach (see paragraph 3.2 on the "Human Simulation Approach"), in order to capitalize on a research field with a long psychological tradition (L1 acquisition), in which the importance of social interaction has been explored with particular emphasis. Indeed, most models of L1 word learning specifically acknowledge the importance of a social partner, whose presence is particularly helpful for the learner because of the attentional value. In particular, this value has been attributed to the fact that the caregiver responds to the learner's needs in a *punctual* and *timely* fashion. This behavioral adaptation is not a prerequisite of child-caregiver interaction; instead, it represents a typical phenomenon in socially interactive settings in which adult members of a dyad consistently tend to temporally coordinate with each other. What function can this social effect have on word learning? It has been proposed that coordination phenomena occurring naturally in dyadic situations may be due to endogenous brain oscillators becoming entrained to each other. This explanation, originally proposed to explain turn-taking, allows several pieces of evidence to be bound together (for example, temporal coordination between caregiver and child, temporal coordination between adult learner and teacher, or coordination of brain activity). Further, this concept provides a minimal account of the effects of social interaction on word learning: The mapping of a word with its referent is critically dependent upon the timing of its presentation, which is ensured by the temporal coordination between members of the dyad as an emergent property of entrained endogenous oscillators. This explanation, which has been proposed for L1 acquisition, constitutes a feasible account easily applicable to the effects of social interaction on L2 learning. However, as previously highlighted, how and whether social interaction influences vocabulary learning in adults is a field yet to be explored.

It is vital to note that the answer to the question “does social interaction influence word learning?” requires a multi-disciplinary investigative approach, drawing on several fields of investigation. On the one hand, a comparison between L1 and L2 may help in identifying similarities and differences in the learning mechanisms, by identifying what has been overlooked in one or the other field (for example, social interaction in L2). However, social interaction represents a branch of psychological and neuroscience research with its own tradition and methods (see chapter 1 and paragraphs 3.1 and 3.2 in the Methods section). For example, the analysis of temporal coordination is typical of joint action studies; nevertheless, as the entrained oscillators theory suggests, its application may be particularly useful for language learning as well. An investigation focusing only on one specific field is at risk of missing important information as to how, for example, different sources of information are integrated. A critical aspect in this regard is represented by the possible integration of local linguistic information and social interaction. As described in paragraph 1.2.1.3, both adults and children are able to learn the meaning of a novel word after just a few presentations, while refinements to the concept occur with several presentations of the verbal label. Whether these repetitions may be more beneficial in a consistent or in a varied context is still a matter of debate. In general, a more consistent context may benefit learning more, by strengthening the already created memory trace in an episodic manner; in contrast, a more varied context may improve learning by enriching the representation of a lexical item. However, it is reasonable to hypothesize that the advantage of one or the other type of encoding may be significantly different in a social or in a non-social situation. Consider the example of the “gavagai” tale, proposed by Quine (Quine, 1960), in which a linguist is exposed to words in an unknown language and has to identify the correct referent in a world of infinite possibilities. The first time the learner hears “gavagai”, an Aruntan speaker is pointing at a rabbit and saying “The gavagai is running”. Let’s assume the learner will infer “gavagai” to mean “rabbit”, in virtue of the whole-object bias (see paragraph 1.2.1.2). The second time the learner encounters the word “gavagai”, it will not really matter if someone is pointing at the rabbit: The learner will infer “gavagai” to mean rabbit consistently with her/his first exposure to the word. Assume, however, that the learner is now in a completely different context, and someone says: “my Gavagai’s name is Pip”. Without someone indicating who or what is Pip, the learner will face the problem of finding the referent all over again. In this case, a consistent context may be more helpful: Since the referent is already known, the subsequent presentations of the word will most likely strengthen the already formed memory trace. However, if a native speaker is there to indicate the referent, a varied context may lead to a more refined memory trace, as with each occurrence new elements are added to the concept depicted by the verbal label while its referent is signaled by the social partner. Hence, a prediction based on this example is that word learning may be better for words repeated in a consistent context in non-social situations, but in interactive situations, it should be improved when repeated in a different context. This example clearly indicates how social interaction may be a source of modulation in adult word learning, while at the same time emphasizing the importance of an approach drawing on multiple sources (linguistics, social psychology, learning and memory research).

While the question of whether social interaction may influence word-learning represents the general rationale motivating the studies presented in this dissertation, each study has focused on a specific facet of this general query. More specifically, the first study explores the possibility that temporal coordination between partners may facilitate word learning. The second study investigates the specific role of social interaction on word learning in comparison with music, a stimulus with similar characteristics (such as complexity, interactivity, and rhythmicity; see paragraph 1.2.2.6) as a coordinative tool. The rationale behind this study is that, if temporal coordination per se drives learning, then there should be no difference in the learning outcome for different stimuli as long as they have similar potential to elicit temporal coordination in the learner. The third study explores the neural correlates of adult word learning, by focusing on the effect of a social partner on the attentional system of the learner. A further aspect investigated in each study is the interplay between local properties of the context (that is, variability in the sentence context) and the social context.

While the specific hypotheses addressed by each study will be highlighted in the correspondent introductory paragraphs, the research questions addressed in this dissertation may be summarized as follows:

1. Is social interaction a factor relevant in second language learning? [*Study 1, Study 2, Study 3]*
2. Does temporal coordination between the learner and social partner drive learning? [*Study 1, Study 2]*
3. Do all external "oscillators" facilitate learning, or is a human partner "special"? [*Study 2]*
4. Does entrainment with an oscillator entail attentional properties? [*Study 1, Study 2, Study 3]*
5. How does social interaction interact with properties of the local context (variability in the sentence context)? [*Study 1, Study 2, Study 3]*
6. Does social interaction modulate brain activity during word learning? *[Study 3]*

3

Methods: An interactive approach to the study of word learning

The previous chapters have defined social cognition and discussed its neural correlates, as well as the mechanisms through which it might influence word learning. Subsequently, the main processes and theories of language learning have been explored for both first and second language acquisition. To complete this introductory part, an overview of the relevant research questions and hypotheses will now be provided.

In this chapter, the historical background on the paradigms developed to study social cognition will be discussed, from the point of view of social interaction. These methods will be evaluated in light of recent discoveries that show how the role played by an observer during social situations may be significantly different from that of someone with an active role in the scene. This evidence calls for a new social paradigm, allowing the study of participants during real interactions, and not merely during observation of social events. Bearing this in mind, a new interactive set-up was developed, and its rationale and characteristics are explained in this chapter. This interactive method was used in the set of experiments described in this thesis for both behavioral and neuroimaging acquisitions, and the last part of this chapter will explain the rationale of using these methods.

3.1 Social paradigms: On the difference between being an observer and being an actor

Are we influenced by others in our daily activities? This question has fascinated scientists since the beginning of modern psychology. The origin of social psychology may indeed be traced back to 1898, when Tripplett discovered that cyclists were faster when racing with a pacemaker than when racing alone (Strube, 2005; Triplett, 1898), a phenomenon later called “social facilitation” (F. H. Allport, 1924). In brief, the “social facilitation” effect states that people tend to perform better in simple, over-learned tasks when they are in the presence of others. However, when the task is difficult, the presence of an audience has the opposite effect and in fact hinders the performance (Bond & Titus, 1983). Importantly, the social facilitation effect is observed when other persons are simply present on the scene and do not intervene with an explicit behavior; indeed, in the latter case, an effect of social facilitation may be difficult to interpret because it would be confounded by other social factors (Zajonc, 1965; Strauss, 2002).

Interest in the complexity of social settings involving interaction persists up to the present day, and stretches from the field of social psychology into social neuroscience research. In this field, the development of paradigms involving real-time interactions was hindered for a long time by both theoretical and technical concerns. On the theoretical side, since the dawn of social neuroscience (Frith & Frith, 1999; Ochsner & Lieberman, 2001), the “social” part of social cognition was confined to the ability to perceive others’ minds (Frith & Frith, 1999, 2006, 2012) or actions (Rizzolatti & Craighero, 2004; Rizzolatti & Fabbri-Destro, 2008). In this scenario, social interaction was considered an ancillary aspect to the more important human ability of understanding each other. This predominant view found fertile ground in neuroimaging studies: Here, theoretical arguments meet complications arising from the sensitivity and limitations of the technical equipment, leading to a preponderance of studies in which the participant is merely an observer of social situations (Schilbach, 2014; Schilbach et al., 2013). This shortcoming holds true even in studies investigating social phenomena requiring social interaction in real life. As an example, learning a new vocabulary by observing social situations is taken as a type of social learning, despite the fact that the learner has a passive role (Jeong et al., 2010).

Recently, however, awareness is growing in the field of social neuroscience regarding the importance of investigating social cognition from the point of view of interaction. This view stems from the widespread idea that the ability to understand others' minds is not all that matters in social exchanges – it is also necessary to consider the reciprocal relation with the person one is interacting with. This type of situation has indeed been proven to strongly influence the behavior of the people involved (Hasson et al., 2012; Knoblich & Sebanz, 2008). As an example, interacting dyads tend to achieve coordination in their behavior, even if not required to do so (Richardson et al., 2007; Shockley et al., 2009). Further support for the importance of social interaction emerges from recent fMRI evidence showing how both areas involved in perceiving another's mind and actions are recruited when someone is required to directly interact with someone else instead of being merely an observer (Ciaramidaro et al., 2014).

Taken together, this evidence contributes to the idea that situations in which a participant simply *sees* a social stimulus may be significantly different from those where the person is truly *interacting* with someone (Jeong et al., 2011; Liu & Pelowski, 2014; Schilbach, 2014); further, it calls for new paradigms and studies focusing on real-time interactions. The paradigm described in the following sections was developed as a response to this rising demand, by allowing real on-line interaction to take place during word learning. Before providing the details of this paradigm, a well-documented setting used to evaluate word learning in healthy adults is described.

3.2 Word learning paradigms: The Human Simulation approach

As described in previous chapters, both children and adults are quite remarkable in their ability to pick up contextual elements, helping them to decipher a new word's meaning. Social, linguistic, and attentional cues are integrated and exploited to map verbal labels to their respective referents. One approach to the study of word learning has been to re-create a similar learning situation, in which the learner has to rely on context to make a first guess as to what a new word means.

The name "human simulation paradigm" to denote such a set-up first appeared in a study by Gillette and colleagues (Gillette et al., 1999) intended to investigate how new words are acquired. More specifically, the authors wanted to investigate the bootstrapping hypothesis, according to which vocabulary acquisition emerges during an incremental process in which linguistic representations acquire progressive levels of complexity (see paragraph 1.2.1.2 for more information). The term "simulation" is derived from the fact that this paradigm aims to simulate an ecological learning situation experienced by children, in which both linguistic and extra-linguistic elements are provided to the learner, but using adult learners. More specifically, the authors asked adult participants to identify in a dialogue target words masked with novel words (such as "flurg" or "glorp") or beeps, under varying informational circumstances. Similar to what happens in children, the adults in the experiments identified significantly more nouns than verbs, and verbs with greater imageability (for example, to run) were recognized more easily than those with low imageability (for example, to think). Hence, the human simulation approach was successful in reproducing results observed in children in a controlled experimental setting.

A similar approach was used by Mestres-Missé (Mestres-Missé et al., 2007); in a series of studies, Mestres-Missé and colleagues investigated the neural correlates of word learning in healthy Spanish adults, by presenting them with triplets of sentences all ending with the same novel word. The participants' task was to extract the meaning of the novel word from the context of each sentence in the triplet (Mestres-Missé et al., 2008, 2007; Mestres-Missé, Rodriguez-Fornells, & Münte, 2010). These studies successfully provided the first evidence that learning from contextual information is a reliable and efficient strategy for adult learners, in which cortical areas are activated in concert with subcortical structures.

The human simulation paradigm thus provides an efficient and ecological way to test word learning in adults under variable conditions of information availability. For this reason, its principles were integrated in the paradigm described in the following section.

3.3 A new paradigm to study word learning during social interaction

The set-up described in this chapter was developed to enable the study of word learning during an “online” interaction between a participant and a more knowledgeable person (an experimenter), by capitalizing on experiences from both social interaction and word learning paradigms. The particular asymmetrical nature of this setting bridges first language learning (Csibra & Gergely, 2009; Gogate, Bahrick, & Watson, 2000; Hirotani et al., 2009; Pereira et al., 2008), and second language learning, where it represents an increasingly common situation. As an example, one can imagine a situation where a young adult travels abroad to learn a foreign language, and the people she/he will meet are experts in their own language, despite not holding a teaching degree. While talking together, the young adult will learn new words by picking up information from both the linguistic context (that is, the context of the sentence in which the word is embedded; Mestres-Missé et al., 2007; Rodríguez-Fornells, Cunillera, Mestres-Missé, & de Diego-Balaguer, 2009) and the behavior of her/his knowledgeable partner (that is, movements, expressions, gesturing; Jeong et al., 2010; Mol, Krahmer, Maes, & Swerts, 2012).

How is it possible to study such a complex scenario within the limits of an experimental setting? Which elements in the context does the learner use? And are all words equally easy to learn? The setting described in the following chapter aimed to answer these questions by recreating an ecological learning context. In the set-up, new word referents and other referents are intermixed, recreating a situation in which the learner has to ascertain the meaning of a new word among several possibilities. The learner is able to find the correct meaning with the help of a person (the experimenter) who has greater experience in the new environment. Similar to what happens in real-life situations, the experienced partner will also provide the sentence context in which the word is embedded. Further, by manipulating the repetitiveness of a sentence context, this setting allows one to test whether the presence of a knowledgeable person is more helpful for simpler compared to more difficult situations, a question that remains unanswered since the beginning of social psychology (T. Straube, Schulz, Geipel, Mentzel, & Miltner, 2008; Strauss, 2002; Triplett, 1898; Zajonc, 1965).

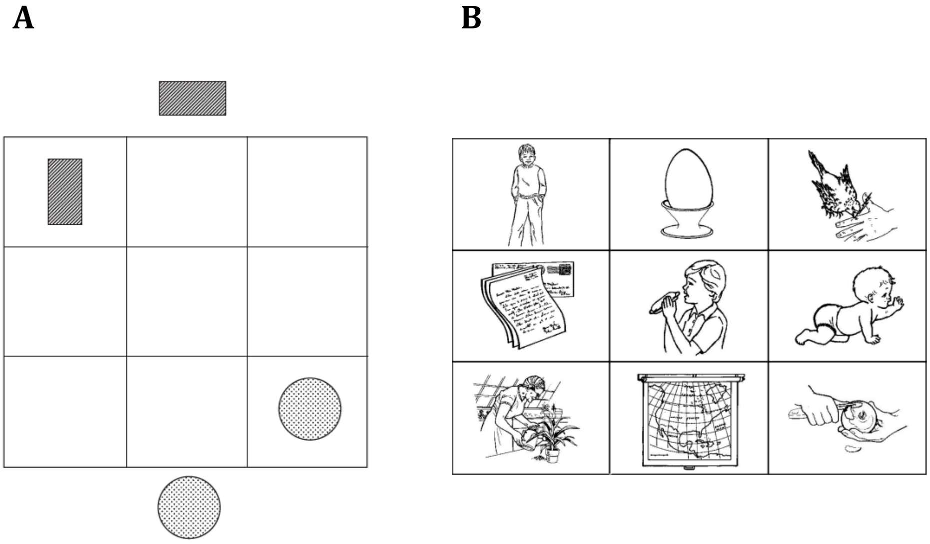
The set-up takes shape as a sort of game; this choice of design is in line with the literature, as game-like tasks are frequently used in social neuroscience (for example see King-Casas et al., 2005; Montague et al., 2002; Redcay et al., 2010). A particularly interesting approach, on which the principles of the learning task presented here are based, is the Tacit Communication Game (De Ruiter et al., 2010; De Ruiter, Noordzij, Newman-Norlund, Hagoort, & Toni, 2007; S. Newman-Norlund et al., 2009; Willems et al., 2010). In said game, pairs of participants are seated behind separate screens displaying a 3 x 3 grid and two geometrical shapes (called "tokens"), each governed by one of the participants. The participants' task is to move the tokens on the grid to reach a final configuration known only by one of the two participants. Therefore, in these trials, a pair may solve the game only if the knowledgeable participant is able to convey the information he has to the other participant (De Ruiter et al., 2007; see **Figure 5A**).

While the general idea is similar, the stimuli and implementation of the interactive word-learning game are strikingly different (**Figure 5B**). In the word-learning game, a participant is paired with an experimenter to interact on a visual grid containing simple pictures depicting nouns or actions (the word referents); their common goal is to combine the pictures in order to create a plausible German sentence, whose object is associated with a verbal label representing its name in a foreign language. The idea behind the game is that a new name is presented to the participant only when the dyad successfully interacts to disclose the hidden sentence. Crucial to the game is the superior knowledge that the experimenter possesses about the grid and its elements: Since she knows in advance which sentence is hidden in each checkerboard, the experimenter is able to direct the attention of the learner towards the relevant elements in the grid (Csibra & Gergely, 2009; Louwerse, Dale, Bard, & Jeuniaux, 2012; Pereira et al., 2008).

This type of set-up offers several advantages, most notably the possibility of extending the study of social interaction to different populations and techniques; indeed, in the studies reported hereafter, healthy adult participants have been tested by means of both behavioral and fMRI approaches. However, the easiness of the task potentially allows this set-up to be employed with more

Figure 5 – Social games.

Examples of stimuli presented in the Tacit Communication Game (A, reproduced from De Ruiter et al., 2007) and in the word learning game (B).



demanding populations, such as children or patients with language disorders. This latter point will be discussed as an overview in the last chapter.

In the following section, the selection of the stimuli is described, as well as details concerning the assembly and balancing of the checkerboards.

3.3.1 Stimuli selection and balancing

The first step in the game's implementation consisted of creating the sentences to hide in the checkerboards. Based on the existing literature, a total number of 25 new words to be learned per subject, repeated 11 times each, was considered to be a good balance between experimental time and learning efficiency. Since the sentences had to be conveyed by images, the first stage was to select the pictures, which were required to be clear, not too complex, simple to recognize and able to unequivocally convey the intended meaning. To fulfill these requirements, a set of black-and-white drawings of objects, humans, animals, and actions was selected from a validated database of pictures available on-line (Bates et al., 2003; Szekely et al., 2003, 2004, 2005; http://crl.ucsd.edu/experiments/ipnp/). This source was chosen because it provides a large number of easily recognizable images for which normative studies and several linguistic measures are available in multiple languages. Besides the crucial advantage of allowing a more controlled balancing of the stimuli, this also increases the possibility of meaningfully exporting the set-up to different languages in future studies.

3.3.1.1 Pictures and sentences

A set of 566 pictures were originally selected and grouped into the following semantic categories: People (N = 33), animals (N = 78), body parts (N = 28), vehicles (N = 33), food (N = 46), things to wear (N = 36), small artifacts (N = 161), large artifacts (N = 75), objects/phenomena in nature (N = 24), and transitive verbs (N = 52). Pictures depicting objects in the plural form (for example, “Handschellen”, *handcuffs*), objects with a compound name (for example, “Briefkasten”, *letter box*), objects representing general categories (for example, “die Gemüse”, *vegetables*) and impersonal verbs (for example, “regnen”, *to rain*) were discarded.

The corresponding names of the remaining pictures were then further screened and ultimately combined to create plausible German sentences, using a stepwise procedure. First, a set of nouns with high naming frequency (according to the CELEX database for German language) was selected. Second, all possible combinations of nouns and verbs composing a plausible German sentence were created. Third, sentences created in the second step were excluded if they had the same subject and object (for example, “Der hund verfolgt den Hund”, *the dog follows the dog*) or if the subject was inanimate (for example, “Der Wind biegt den Baum”, *the wind bends the tree*). Fourth, nouns were divided into two groups similar in composition for naming frequency, gender and animacy, and one of the groups was assigned to “subjects” and the other group to “objects”. The group “objects” was further divided in two balanced groups: Objects for sentences whose context was the same at each repetition (same Same Context, sSC), and objects for sentences whose context was different at each repetition (Different Sentence Context, dSC); these categories were used to evaluate whether the presence of a knowledgeable person is more helpful for simpler learning contexts (the repeated sentences) compared to more difficult (the variable sentences; T. Straube et al., 2008; Strauss, 2002; Triplett, 1898; Zajonc, 1965). The sentences, at this stage, constituted a pool of basic “template sentences”, of which elements in the subject and object positions could be easily replaced; for example, in the sentence “*the woman cuts the bread*”, the subject “*the woman*” can be easily substituted for other human subjects, such as “*the man*”, “*the boy*”, and so on. Given this property, it was possible to balance the sentences for a set of parameters including semantic category, gender, and animacy. The

Figure 6 – Action picture modifications.

Example of an original action picture containing an object (left) and its modified version (right). The action depicted in the figure is "*to peck*". The object was modified to become unrecognizable, in order to avoid confusion or priming effects (for example, the participant was referred to "*hand*" as the object of the sentence).

final set included 394 sentences (300 dSC and 76 sSC[12]), balanced for naming frequency, animacy, semantic category and gender.

Action pictures (verbs) including objects were modified when possible to avoid priming or interference effects; for example, if the image representing the verb "to peck" included a bird pecking a human hand, the hand was graphically modified not to be recognizable to avoid participants being primed or confused by the hand. This example is depicted in **Figure 6**. Modifications were made with the deformation option implemented in the graphic suite GIMP 2.6.8 (www.gimp.org); all modifications were conducted with a warping procedure performed using a deformation radius of 15 mm and a deformation amount of 100 mm.

The sentences we created were subsequently evaluated in two studies, both in their written form (Validation Study 1) as well as in their pictorial form (Validation Study 2). In the first study, sentences were evaluated for their

---

[12] We only required 25 sSC sentences, corresponding to the number of words to be learned, since each of them was then repeated 12 times. The final pool of 25 sentences was selected after the results of the two evaluation studies.

plausibility and how well the object fitted in the context of the sentence. In the second study, a judgment of sentence plausibility was performed on the pictorial version of the sentences; moreover, participants were required to write the sentence that was depicted. This way, it was also possible to ensure that the modification of the action pictures did not interfere with the intelligibility of the sentence. Details of the validation studies are provided in paragraph 4.1 and 4.2.

3.3.1.2 Pseudo-words

The aim of the game was to teach new words to cohorts of native German speakers. The Italian language was chosen since it is not a language frequently taught in German schools, compared to English or French. However, to ensure that participants had not previously encountered the words incidentally, Italian pseudo-words were adopted to represent the learning target. A set of pseudo-words was extracted from a published paper investigating the role played by Broca's area in speech perception in relation to the lexicality of a stimulus, using Italian words and pseudo-words (Kotz et al., 2010). The original set of pseudo-words comprised 161 disyllabic pronounceable pseudo-words (for example, "*tasna*"; length range: min. 4 max. 6 letters). From this pool, an original sample of 50 letter strings was selected to comply with the following requirements: First, pseudo-words were required to end with the vowel "a" or "o"; this was chosen to avoid confusion with the typical plural form of Italian words (usually "i" or "e"), since all of the pictures depicted nouns in the singular form. Second, they were not cognates of existing German words. Third, they "sounded Italian" to German native speakers (N = 5). The final sample of 50 words was then rated by 15 Italian native speakers, who were required to indicate a) whether the words were plausible in terms of Italian word structure, b) the possible origin of the words, and c) how difficult it was to retrieve the original word. All the pseudo-words were rated above 3 on a Likert scale from 1 to 5. As the pseudo-words had previously been used in an experiment with Italian participants, this rating should be considered as an additional quality check on the material.

The final sample of pseudo-words was then divided into two groups to be assigned to the sSC (N = 25) or dSC (N = 25) categories, and balanced both within and between groups for syllabic complexity, last letter (ending with “a” and “o”) and first letter.

3.3.2 Checkerboards: Preparation and balancing

The last step in the implementation of the game consisted of creating the checkerboards, starting from the plausible sentences created from Validation Study 1 & 2.

In this phase, the sentences - in their pictorial form - were positioned on the checkerboards. The three pictures were organized in the grid so that they were touching each other at least corner to corner. Each checkerboard was created manually, to ensure that each of the elements in the sentence appeared in each position on the grid a comparable number of times. All possible combinations of positions on the grid were employed a comparable number of times. The remaining six pictures in each checkerboard (excluding those belonging to the sentence) were distractor images chosen from the initial image pool and were balanced between pictures representing nouns (either animals, humans, or objects) and actions. These distractors were selected to ensure that none of them could be considered as an additional plausible object for a given sentence context; this way, given the constraint that the elements of the hidden sentence needed to be positioned close to each other, only one object in each checkerboard could represent the correct target object. The checkerboards were further balanced in terms of the mean naming frequency of the items depicted by the 9 pictures both within and between checkerboards.

3.4 Behavioral measures

How do we know if someone is learning? Usually, a skilled person can be distinguished from a trainee because of his *faster* and more *accurate* behavior when performing the same task. These two parameters – speed and accuracy - are critical in the description of the learning process. Usually, the speed of responses in a given task (often termed *response time* or *reaction time*) increases with learning (Lemmon, 1928). In parallel, the better we learn to perform a task, the more accurate we become. The word-learning game described in the previous paragraphs allowed us to evaluate both response times and accuracy; more specifically, *response times* were calculated as the delay between the presentation of the sentence context and the selection of the correct object by the participant, while *accuracy* was defined as the proportion of correct responses given by the participant during the learning phase.

The main focus of this dissertation is social interaction and, as previously stated, social interaction is related to *anticipating* and *adapting* to someone else; indeed, it has been demonstrated that temporal behavior during social interaction becomes coordinated (Richardson et al., 2007). To evaluate this behavior during word learning, an additional set of measures was considered. First, the standard deviations of response times were employed as a measure of participants' performance, and second, correlations between the response times of the experimenter and participant were calculated to measure the degree of coordination inside the dyad. These types of measures are typical of studies evaluating the behavior of participants in dyadic settings (Kawasaki et al., 2013; Louwerse et al., 2012; Pereira et al., 2008) and in settings requiring coordination with an external source (Repp, 2005; Repp & Su, 2013; Vesper, van der Wel, Knoblich, & Sebanz, 2011).

Lastly, the end product of the learning process consists of creating a trace so that it is possible to retrieve the learned item from memory; in the case of the word-learning game, said item is represented by the words presented during the learning phase. To ensure that said words had indeed been learned, a testing phase was conducted in each of the experiments immediately after the learning phase. *Accuracy* and *response times* were calculated for each participant in this phase to provide a characterization of the learning outcome. The implementation of the testing phase is specified in chapter 3 for each of the presented studies.

3.5 Neuroimaging

While behavioral measures provide important information concerning the evolvement of a learning process, they leave the question of what happens in the brain during learning open. Neuroimaging techniques are frequently employed as a method to move inside the "black box" that is the brain. Which areas are involved in word learning? How are these areas connected together? And how are these circuits modulated by specific conditions? These are some of the questions that it is possible to answer using a specific neuroimaging technique, namely magnetic resonance imaging (MRI; Huettel, Song, & McCarthy, 2004).

MRI is the method of choice whenever the focus of an investigation involves the brain, particularly intended as specific areas and their connections. Indeed, its spatial resolution in the order of millimeters allows the creation of high-resolution images of the brain. MRI is based on the property of proton nuclei to rotate around their own axis (or *spin precession*). When a magnetic field is applied, the proton spin tends to align to it either in a parallel or anti-parallel direction (*longitudinal magnetization*) and to precess with a specific frequency (*Larmor Frequency*). If a radiofrequency pulse (*RF*) is applied, the proton spin will tend to drift away from the longitudinal magnetization and eventually reach a new position (*transversal magnetization*) in which all spins are equalized with respect to the original parallel or anti-parallel alignment. The angle between the direction of the transversal and longitudinal magnetization is called *flip angle*. In the new configuration, the neutron spins start to interact with each other and move out of phase: This leads to an exponential decay from transversal back to longitudinal magnetization, called *relaxation*. The relaxation time depends on the type of tissue the neutrons belong to (bone, cerebrospinal fluid, white matter, and grey matter all have different relaxation times); the time required to return to longitudinal relaxation provides an indication of the tissue from which the signal originates. Indeed, different types of relaxations (for example, T1, T2, and T2*) are used to weight the contrast that will be used to create the brain images differently; for example, T1 images are conventionally used for anatomical images to represent a good contrast between white matter (white) and grey matter (grey).

Even more interestingly, MRI not only allows the study of brain structure, but also of the brain's activity, by capitalizing on the amount of oxygen present in the blood flowing in brain vessels. This particular type of MRI is called functional magnetic resonance imaging (fMRI). Halfway between legend and history, the discovery of the principles of fMRI trace back to Angelo Mosso, the Italian physiologist who discovered that blood flow towards the brain increases during mental activities, such as performing mathematical operations (Sandrone et al., 2014). This observation paved the way for the development, several years later, of the Blood Oxygenation Level Dependent (BOLD) fMRI (Ogawa & Sung, 2007); in brief, this technique capitalizes on the different iron content in oxygenated versus deoxygenated blood. When the brain performs a certain activity, more oxygen is consumed in the region involved in the task, and this determines an increase in the deoxy- over oxyhaemoglobin ratio. As oxygenated haemoglobin is diamagnetic, while deoxyhaemoglobin is paramagnetic, this in turn elicits a perturbation in the local magnetic field, given the paramagnetic properties of deoxyhaemoglobin (Matthews & Jezzard, 2004). This perturbation speeds up the T2* relaxation, leading to a signal loss; however, blood flood quickly compensates for the use of oxygen, so that the level of oxygenated blood is higher in the activated region compared to the rest of the brain, leading to an increase in relaxation time which is interpreted in the subsequent image *reconstruction* as an activation of that area (Huettel et al., 2004; Matthews & Jezzard, 2004; Ogawa & Sung, 2007).

In this way, fMRI is able to provide a measure (although indirect) of how much an area is recruited for a given task. Before this information can be extracted, however, raw functional data need to be *pre-processed*. More specifically, *slice timing correction* is applied to correct for time differences due to the fact that not all slices in a brain volume are acquired at the same time. Afterward, all the images are *realigned* to a reference volume (usually the first volume of the first session) by means of a rigid body transformation, to account for displacements caused by movements during scanning. An additional improvement to the images can be carried out at this stage by applying pre-calculated *Voxel Displacement Maps* (VDM), to remove artifacts due to inhomogeneity in the magnetic field. The images of each participant are then *co-registered* to the high-resolution anatomical scan of the same participant, *segmented* into grey matter, white matter and cerebro-spinal fluid and finally *normalized* to a standard anatomical template[13] to facilitate inference on a

---

[13] One of the most common templates is the Montreal Neurological Institute (MNI) template.

group level. Lastly, data are smoothed to suppress noise and residual inter-individual differences[14].

The pre-processed data are then statistically analyzed with a two-stage approach: First, the same General Linear Model (GLM) is applied singularly to each participant and the relevant parameters are estimated; these estimates are then fed into a second-level GLM, providing the Statistical Parametric Maps (SPMs) at the group level. The type of statistical approach to be used is dependent on the research question; details on the analysis used for the study described in this dissertation will be provided in chapter 3.

However, this is not all that functional magnetic resonance has to offer. Indeed, new and more refined methods have been developed to extend the principle of fMRI to the study of brain networks, rather than single areas (Penny, Friston, Ashburner, Kiebel, & Nichols, 2011). Some of these methods, for example, use the correlation of functional signal to identify a set of areas (sometimes termed *components*) coherently activated together (Independent Component Analysis – ICA; Calhoun, Adali, McGinty, et al., 2001; Calhoun, Adali, Pearlson, & Pekar, 2001). Even more interestingly, methods of functional connectivity nowadays allow the study of connections between brain areas, further allowing the experimenter to test specific models based on empirical and theoretical evidence (for example, Dynamic Causal Modelling – DCM; Friston, Harrison, & Penny, 2003; Marreiros, Kiebel, & Friston, 2008; Stephan et al., 2007, 2010) . These methods are described in more detail in the following section.

---

[14] The order in which the pre-processing steps are performed is not entirely fixed; while the described pipeline is one of the most common, the order of the steps may vary according to the specifics of the study at hand.

3.5.1 Independent Component Analysis (ICA)

Independent Component Analysis (Bell & Sejnowski, 1995) aims to separate the resulting multivariate signal into independent non-gaussian signals. A classical example of ICA application is the case of the so-called “Cocktail Party effect”, namely the ability to isolate a specific auditory stimulus (for example, the voice of a friend speaking) from a multitude of other voices.

The case of fMRI is somewhat less intuitive; as pointed out in the previous section, the BOLD signal is an indirect measure of how much an area is recruited in a given task, and this means that several temporally and spatially overlapping sources add up to create the signal. ICA is able to separate “noise” components, in which the signal is driven by physiological and artifact related signal sources (Kiviniemi, Kantola, Jauhiainen, Hyvärinen, & Tervonen, 2003). This is not the only clear advantage provided by ICA; indeed, this method has important applications in the study of connectivity of brain areas, as it is able to separate the BOLD signal into independent spatial maps, enabling the analysis of co-activation in spatially divergent areas within a given map (McKeown & Sejnowski, 1998). This is not possible with standard univariate approaches even when two areas are activated by the same task; although this would suggest that activation in the two areas correlate with the experimental manipulation, it does not necessarily mean that activity in the two areas is correlated. It is important to note that ICA is a method to *blindly* separate the signal subcomponents, and it does not rely on a-priori assumptions. For this reason, it is often used for exploratory analysis of fMRI data (McKeown et al., 1997).

3.5.2 Dynamic Causal Modelling (DCM)

Dynamic Causal Modelling (DCM) is a method used to investigate effective connectivity, which allows the inference of hidden neuronal states from measurements of brain activity (Friston et al., 2003). While effective connectivity is often contrasted with functional connectivity, both families of methods in fact use measures based on the BOLD signal; as such, they are different from other measures investigating structural connectivity through white matter assessment (Mori & Zhang, 2006). However, it must be pointed out that the distinction between structural and functional/effective measures of connectivity is mostly methodological. Indeed, while based on different measures, structural and functional/effective connectivity are not independent and reflect two faces of the same coin (Kahan & Foltynie, 2013; Stephan, Tittgemeyer, Knösche, Moran, & Friston, 2009).

While both deal with measures of the BOLD signal, the difference between functional and effective connectivity methods lies in the ability to specify causal relations between distal regions: Whereas functional connectivity methods only describe correlations between brain activity in different areas, effective connectivity methods consider the direct influence that a region exerts on another (Kahan & Foltynie, 2013; Stephan et al., 2010). In other terms, effective connectivity is *directional*, as it allows the definition not only of areas of interest, but also of their connection and respective directions. Methods of effective connectivity, such as DCM, are incredibly powerful, as long as they are based on strong a-priori hypotheses.

A typical DCM model comprises two different levels: The first level consists of the measured time series (for example, the BOLD signal), while the second level is represented by the underlying neural dynamics, which are hidden from direct observation and do not correspond to any common neuropsychological measure. The function of DCM is to model how the neural dynamics are influenced by external inputs and to reconcile the observed BOLD signal with this hidden level. As for the first point, externally induced perturbations – such as those induced by experimental manipulations – may a) directly influence the activity of specific regions (for example, evoked responses in visual areas), or b) modulate the strength of coupling among regions (for example, learning or attention). It is important to note that the distinction between “driving” (or direct) and “modulatory” inputs represents a neurobiologically plausible account, where synaptic responses in the target region are directly driven by inputs, or indirectly driven by inputs from another area. As for the second point, DCM combines this model of neural dynamics with a biophysically plausible and experimentally validated haemodynamic model that describes the transformation of neural activity into the BOLD response; this is called the “Balloon Model” (Buxton, Uludağ, Dubowitz, & Liu, 2004; Buxton, Wong, & Frank, 1998), and allows identification of the neural and haemodynamic parameters from the BOLD measures. Once the model has been calculated for the haemodynamic data, the posterior distributions of the parameters can be used to make inferences on effects at the neural level.

As previously noted, DCM is an a-priori-based approach, in which areas and their connections have to be specified in advance. For this reason, the method relies on the definition of a set of a-priori areas of interest. To create a strong model, the areas should be defined a-priori based on the literature, and present in the second-level GLM of the study in which they are to be used. From these areas, local maxima for each subject should be selected to define specific volumes of interest (VOIs) from which a second-level DCM model can be defined at the group level, to finally estimate the distribution of the posterior parameters.

4

Studies: Social interaction in second language word learning

In this chapter, five studies are presented. Given the novelty of the set-up we employed, the first two studies were conducted to validate the material.

Each of the following main studies addresses a specific question concerning social interaction in second language word learning. Are we self-sufficient learners, or are we – to some extent – like children? What happens when we learn with someone else? Are there only changes in behavior, or does it also influence the way our brain processes the information? These are some of the questions this thesis tries to address. Answers to these questions would not only significantly deepen our knowledge of the learning process in adulthood, but may also provide significant suggestions about how to improve learning methods used daily in real life.

The third study deals with the most fundamental question: Is learning with someone else different than learning alone? The rationale behind this study refers to a parallel with first language learning, as even in this context, an initial challenge faced by infants is to build up a vocabulary. To achieve this goal, the presence of a knowledgeable caregiver is pivotal in guiding the child's attention to the correct referent for a new word. This guidance of attention is, in turn, dependent upon the establishment of a temporal coordination in the dyad. Even in adults, social contexts often result in the establishment of temporal coordination between communicating partners. Might temporal coordination represent a mechanism underlying social word learning in adults? And how does this mechanism deal with other contextual elements? To address this question, the word-learning game was employed in two cohorts of participants; one group performed the task with an experimenter, while a second group performed the task with a computer. This set-up allowed us to tackle several open questions: First, whether the word-learning game was an appropriate learning device; second, whether temporal coordination was achieved in this task during social interaction; and third, whether this temporal coordination (and hence social interaction) would enhance word learning in adults. The answer to the first two questions was positive and confirmed our hypotheses that the set-up was highly enjoyable for participants and led to a significant number of words learned without stressful effort. Further, social-learning participants temporally coordinated their behavior to that of a partner. The third question resulted in a more complex answer: increased coordination during social interaction per se did not affect recall or recognition of words in the testing phase. Rather, learning was influenced by a combined effect of temporal coordination and sentence context variability; the more difficult the task, the more social interaction was used to accomplish it. These results provide the first evidence that social interaction plays a role in adult word learning. However, adults are more skilled learners than children; that is, they do not require “help” for easy tasks, but they do when things get more complicated.

The next study investigates the specificity of this social effect. From the first study, it emerges that temporal adult learners might use coordination as a guidance of attention during the word-learning game. Nevertheless, is the presence of another person necessary? Temporal coordination might be achieved with other external pacers, which might be even more reliable than a human being. Hence, a reliable external pacer might enhance coordination – and, in turn, learning – to an event greater extent than a social partner. Or is a social pacer somehow “special”? To answer this question, in the second study we compared social interaction with another complex stimulus known for its ability to induce coordination – music. Simple novel melodies accompanied the presentation of the sentence context in the word-learning game, either in combination with social interaction or without. The rationale for this set-up was to provide the participant with a temporal scaffolding to which he could synchronize. We predicted that spontaneous temporal coordination either with a partner or with music would be better than with a computer. As for the condition in which music and social interaction were both present, worse temporal coordination was expected when compared to the two single conditions; indeed, in this case, the participant would have two different pacers to which she/he could synchronize. This would create confusion as to which of the pacers should be privileged. Again, if coordination drives attention, then higher recognition rates during a testing phase would be expected to correlate with higher coordination during learning. Results replicate the data from the previous study where social interaction is concerned; as for music, temporal correlation was predominant at delayed stages instead of online during the task, and, in general, required more time to be achieved. In other words, the type of coordination achieved with the two pacers was slightly different in nature; this explains why, in the concurrent social and music condition, the two factors competed with each other, hindering coordination. In a subsequent testing phase, we only observed a difference due to the type of sentence context that words were originally embedded in, but no correlations with temporal coordination during learning. While providing potential evidence in favor of a low-level mechanism that supports learning in healthy adults, it was not possible to prove a direct effect on immediate learning. Further studies are required to investigate this point and possible effects on long-term memory.

In the last study, the neural substrates underlying word learning during social interaction were explored. The rationale behind this study was to explore whether social interaction during learning modifies not only our behavior, but also our brain activity. If social interaction enhances the baseline activation in our brain, then the activity of areas involved in the task should be increased when interacting with someone. This would not only mean activity in areas involved in word learning, but also circuits underlying more general purpose functions such as attention. Further, given the results of the previous studies, this should also have a differential influence according to the task demands; more specifically, we expect the effect of social interaction to be maximal when needed more, that is, when the task requirements are more challenging. Results confirmed our hypotheses: Social interaction influenced the activity of areas involved in the task not only at high levels, but also at low levels of analysis (for example, primary visual areas), suggesting a heightened basal activity. More specifically, the activity of high-level task-related areas (for example, the angular gyrus, which is involved in spatial attention) modulates primary areas involved in the first stages of the task (for example, primary visual cortex).

Taken together, the results of these three studies provide a significant advance in the understanding of the mechanisms and substrates underlying word learning in social contexts. More specifically, they suggest that social interaction might have an impact in adult word learning in concert with the characteristics of the sentence context a new word is embedded in.

In chapter 5, the results of these studies will be set in a more general background and integrated in a model of how social interaction may impact adult word learning. In addition, possible clinical implications and subsequent future directions will be provided in an overview.

4.1 Validation Study 1

The first validation study was conducted to ensure that the created sentences represented plausible German expressions. To this aim, 20 native German speakers (9F, mean age 24.50 years, SD 2.74) were presented with 564 written transitive sentences in the form subject-transitive verb-object. The participants' task was to make two judgments on each sentence. The first judgment concerned the plausibility of the sentences; participants were required to express whether they thought the stimuli represented plausible German sentences, by rating them on a scale ranging from 1 (the sentence is absolutely not plausible) to 5 (the sentence is absolutely plausible). For example, the sentence "Das Kind isst das Ei" (*the child eats the egg*) represents a perfectly plausible sentence, while the sentence "Der Apfel isst den Hund" (*the apple eats the dog*) is an example of a non-plausible sentence.

The second judgment required participants to indicate how well the object fitted the sentence context defined by the subject and verb. For each sentence, participants were required to rate, on a scale from 1 (the object does not fit at all) to 5 (the object fits perfectly), whether the object represented a good ending for the excerpt.

To make sure participants were keeping their attention on the task, the sentences were inter-mixed with 188 filler sentences. These filler sentences were grammatically correct but clearly not plausible (for example, "Die Frau zerreisst den Dinosaurier", *the woman tears the dinosaur*). The results of this first study are summarized in **Table 1**.

Table 1 - Stimuli ratings, Validation Study 1.

The table summarizes the results of the ratings provided in the first validation study. sSC and dSC sentences and objects had mean ratings above 3 in a scale from 1 (not plausible / not fitting) to 5 (perfectly plausible / perfectly fitting). The exact number of sentences and objects falling in each rating is provided. Abbreviations: sSC = Same Sentence Context; dSC = Different Sentence Context.

Sentence type	Mean rating	< 3	3-4	4-5
Plausibility				
dSC	4.33 ± 0.61	16	50	234
sSC	4.21 ± 0.70	6	13	57
Filler	1.57 ± 0.52	188	–	–
Object Fitting				
dSC	3.95 ± 0.83	50	92	163
sSC	3.83 ± 0.91	15	24	37
Filler	1.38 ± 0.36	188	–	–

4.2 Validation Study 2

In the second study, the pool of sentences was tested in pictorial form. Participants were presented with three pictures for each sentence representing the subject, verb, and object of the sentence itself. The aim of this validation study was to ensure that the pictures represented the intended meaning and, as a consequence, the plausibility of the sentences as verified in the first study.

20 native German speakers (11F, mean age 24.10 years, SD 2.94) were presented with 376 stimuli, each comprising three pictures representing the subject, verb and object of the sentences that had been judged as plausible in the first validation study (**Figure 7**). In this case we only evaluated the sSC and

dSC sentences, without adding fillers. For each pictorial sentence, participants were required to a) indicate whether the sentence conveyed by the pictures was a plausible one, b) indicate whether the image sequence represented the sentence in a good way, and c) write the sentence they thought was represented by the picture sequence, by typing the sentence into a computer.

Judgments b) and c) were given on the same five-point scale as in the previous experiment, ranging from 1 (the sentence is absolutely not plausible / the pictures are a bad representation of the sentence) to 5 (the sentence is perfectly plausible / the pictures represent the sentence perfectly). Scores for tasks b) and c) were evaluated with the same criteria as Validation Study 1. Further, for each image the agreement between the intended meaning of the picture and the participant's answer was calculated (for example, how many times "*dog*" was given as the answer for the image representing a dog). Images with scores below 3, or with an agreement rate below 50% (that is, the answer "*dog*" was given by less than half of the participants) were not considered further, unless the chosen verbal label was a close synonym of the target word (for example, in German, "Arzt" and "Doktor" both correspond to the English "*Doctor*").

Results are summarized in **Table 2**.

Figure 7 – Example stimuli.

Top row: presented stimuli; bottom row: expected German meanings conveyed by the pictures, and their English translation. Only the pictures were presented to participants.

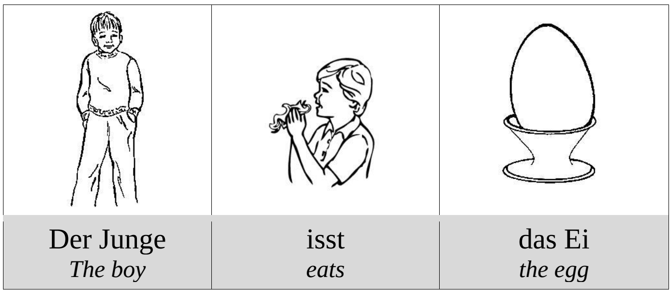


Table 2 - Stimuli ratings, Validation Study 2.

The table summarizes the results of the ratings provided in the first validation study. sSC and dSC sentences had mean ratings above 3 on a scale from 1 (not plausible / not fitting) to 5 (perfectly plausible / perfectly fitting). The mean percentage of agreement between intended and reported meaning is also provided. Abbreviations: sSC = Same Sentence Context; dSC = Different Sentence Context.

Sentence type	Mean rating plausibility	Mean rating object fit	% agreement
dSC	3.81 ± .81	4.35 ± .39	72 %
sSC	3.74 ± .91	4.28 ± .41	66 %

4.3 Learning together or learning alone? Social interaction in adult word learning[15]

Abstract

A first challenge when learning a new language is to build up a vocabulary. In both infants and adults, this form of learning often occurs in social contexts, in which communicating partners temporally coordinate with each other. In children, such temporal coordination facilitates the focusing of attention on relevant aspects in a conversation, which, in the case of word learning, are the possible referents of a new verbal label. Whether similar mechanisms are at work in adult learners during social interaction is the focus of this study. We predicted that participants performing a contextual learning task would temporally coordinate more when interacting with a partner than with a computer. Furthermore, we expected word learning to be better with maximal temporal coordination. German native speakers were exposed to new words in a contextual learning task either with a partner or a computer. Results confirm that participants learning in a social context temporally coordinated their behavior to that of a partner. However, this increased coordination per se did not affect immediate recall or recognition of words in the testing phase. Rather, learning was influenced by a combined effect of temporal coordination and sentence context variability. Taken together, these results provide first evidence that social interaction plays a role in adult word learning. More specifically, the temporal coordination in social interactions modulates word learning in concert with contextual cues. These results highlight the notion that social adult word

---

[15] This study is based on the article: Verga, L. & Kotz, S.A. (in prep.) Learning together or learning alone? Social interaction in adult word learning.

learning is a complex phenomenon in which basic temporal properties may play a significant role.

4.3.1 Introduction

Learning a new language is a complex task that an increasing number of adult learners are currently facing in our modern multilingual world. A first step to succeed in this challenge is to assign meaning to a new verbal label. Despite the apparent simplicity of this task, every new word has multiple referents that need to be defined by cues derived from the context a word is perceived in. In everyday life many of these cues are provided by another person; indeed, most often language learning contexts are also social contexts in which a more knowledgeable person supports the learner in her/his efforts to acquire new words. This is the case not only for infants learning their first language (Csibra & Gergely, 2009; Kuhl, 2007; Kuhl, Tsao, & Liu, 2003), but also for adults acquiring a second language (Jeong et al., 2010; Verga & Kotz, 2013). Frequently, learning occurs in teacher-student interactions in which one member of the dyad has a clear and leading role, even though it is possible for adult learners to acquire new words incidentally from context without explicit training (Laufer & Hulstijn, 2001; Nagy, Anderson, & Herman, 1987; Rodríguez-Fornells, Cunillera, Mestres-Missé, & de Diego-Balaguer, 2009; Swanborn & De Glopper, 1999).

Cues that can be extracted in such social contexts are substantially different from other cues utilized in other types of word learning (e.g., learning from a text book) as a communicative partner delivers them. In this situation, it is crucial for the learner to understand the communicative intention of a partner (Verga & Kotz, 2013). Indeed, the human ability to infer other people's mental state (Theory of Mind or mentalizing) underlies the efficacy of communication between partners by creating a psychological common ground in which information can be successfully exchanged (Frith & Frith, 2006, 2012). On this common ground, learning is facilitated by the fact that partners may reciprocally influence each other to achieve a shared goal. For example, in an asymmetric word learning setting the more experienced partner adapts her/his behavior to guide the learner's attention towards the referent of a new verbal label that has to be learned (Csibra & Gergely, 2009; Newman-Norlund, Noordzij, Meulenbroek, & Bekkering, 2007). This guided focusing of attention dramatically reduces the number of possible referents for the new verbal label as compared to a situation where the learner's attention is not directed towards the correct referent (Louwerse, Dale, Bard, & Jeuniaux, 2012; Rader & Zukow-Goldring, 2012).

In this scenario, temporal coordination between partners is pivotal to successfully triangulate attention between the learner, the caregiver, and the new word referent. Indeed, in order to bind the new verbal label to its meaning, the learner's attention needs to be guided to the correct referent at the correct point in time. The importance of a simultaneous presentation of new words with their referents has been consistently emphasized in first language acquisition (Gogate, Bahrick, & Watson, 2000; Gogate, Walker-Andrews, & Bahrick, 2001; Rader & Zukow-Goldring, 2012; Rolf, Hanheide, & Rohlfing, 2009) together with the importance of the caregiver in directing the attention of the child during this process (Csibra & Gergely, 2009; Kuhl, 2007). For example, in a study by Pereira and colleagues, a toddler-parent dyad was monitored over a period of a few minutes of unconstrained playing in which toy names were presented to the toddler for the first time. The authors measured the temporal coordination between members of each dyad as the temporal coordination between the head and hand movement of the parent and that of the child. Crucially, toddlers interacting in more coordinated dyads performed better in the subsequent word recognition test than toddlers who played in dyads, in which members were both either moving or not moving at the same time; in other words, the turn taking quality was essential for learning (Pereira, Smith,

& Yu, 2008). The authors suggested that a better “social rhythm” (in other words, good temporal coordination between partners leading to smooth turn taking) may support the correct timing of attention to a target. This idea is in line with more general evidence that multimodal rhythms can support the allocation of attention and the integration of information, pivotal to successful learning of new information (Lagarde & Kelso, 2006; Rolf et al., 2009; Schmidt-Kassow, Heinemann, Abel, & Kaiser, 2013).

Despite this first language acquisition evidence, to the best of our knowledge the impact of temporal coordination between partners in second language learning has not yet been investigated. This paucity of research in the field is particularly surprising considering that adults consistently and often unintentionally coordinate during social interactions (Hasson, Ghazanfar, Galantucci, Garrod, & Keysers, 2012; Schmidt, Fitzpatrick, Caron, & Mergeche, 2011; Yun, Watanabe, & Shimojo, 2012). Such temporal coordination emerges at multiple levels during social interaction: Partners tend to fine-tune the timing of their actions not only in ‘simple’ motoric tasks (finger tapping, Pecenka, Engel, & Keller, 2013; jumping, Vesper, van der Wel, Knoblich, & Sebanz, 2012; performing martial arts, Schmidt, Fitzpatrick, Caron, & Mergeche, 2011), but also during higher level cognitive performance such as conversations (Shockley, Richardson, & Dale, 2009). In the latter case, temporal coordination between the speakers is pivotal to establish smooth turn taking, which in turn is essential for effective communication: Both overlaps and long silences should be minimized (Stivers et al., 2009).

The question thus arises whether—similarly to what has been proposed in infant word learning—such pervasive temporal coordination between partners may influence social word learning in adults. To answer this question we created a game-like task in which participants learned new words either with an experienced partner or with a computer. We hypothesized that participants would temporally coordinate more with the partner than with the computer during the learning phase of the task. In turn, we expected that this increased temporal coordination would guide the allocation of attention to the target referent of the new word and ultimately facilitate its binding with the corresponding meaning.

In order to improve the ecological validity of this set-up, we implemented one of the most common word learning situations for adults: Learning from context. As pointed out, adults can incidentally and effortlessly learn new words from context, for example, during conversations. In this particular situation, learning is facilitated by the fact that new words are rarely presented in isolation; more often they are embedded in sentences (Laufer & Hulstijn, 2001; Nagy et al., 1987; Rodríguez-Fornells et al., 2009; Swanborn & De Glopper, 1999) providing helpful cues to disentangle a correct referent (Mestres-Missé, Rodriguez-Fornells, & Münte, 2007; Mestres-Missé, Münte, & Rodriguez-Fornells, 2009). In this case, identification of the correct meaning of a new word may be more or less difficult dependent on how vague the context is (Borovsky, Kutas, & Elman, 2010). If a context does not ensure the recognition of the correct word meaning at first glance, a pivotal role is then played by contextual variability (Glenberg, 1976, 1979): Every time the same word is encountered, different contextual cues accumulate, ultimately facilitating the identification of the target referent and its association with the new word (Adelman, Brown, & Quesada, 2006; Lohnas, Polyn, & Kahana, 2011; Verkoeijen, Rikers, & Schmidt, 2004). Consequently, new words embedded in a consistent sentence context should elicit a faster identification of the correct referent compared to new words embedded in a more variable context. Indeed, if a new word is always repeated in the same sentence context, the choice of the target word only needs to be replicated across repetitions after the first successful identification. Nevertheless, the mapping between words repeated in a more variable sentence context and their referents should be better: each time the word is encountered, more cues will be available from a new sentence context leading to an enriched representation of the word meaning (Adelman et al., 2006; Verkoeijen et al., 2004). Using either variable or more consistent sentence contexts, we therefore aimed to replicate these well-known effects. However, using variable and constant contextual learning also allows us to gain valuable insight into the principles underlying social learning. Indeed, in a social learning situation the learner is exposed to at least two types of cues: the ones derived from the social partner and those extracted from the sentence context. This study is thus intended to provide first evidence on the way these types of cues are integrated. Specifically, we posed the following hypotheses:

In the social version of the task, temporal coordination between partners should be higher for words repeated in variable contexts. In this case, the participant has to decode the sentence context to extract cues directing them towards the correct referent repeatedly. Cues derived from the social partner should help in directing the attention towards the correct referent thus facilitating its identification. In contrast, words repeated in the same sentence context do not require the experimenter to guide the learner's attention towards the target, as the same target will be known from previous presentations in the same sentence context.

Consequently, a variable sentence context in social learning should maximally benefit word learning, resulting in better performance in the testing phase following the learning phase.

If, however, social and context cues are not integrated but are instead used independently, we should expect either a) an increased word learning rate in social context independent of sentence context variability — if social cues are predominant — or b) an increased word learning rate in variable sentence context — if this cue source is predominant.

4.3.2. Methods

4.3.2.1 Participants

Sixty-eight participants took part in the experiment (34F, mean age 25.19 years, SD 2.88 years). All were native speakers of German recruited from the database of the Max Planck Institute for Human Cognitive and Brain Sciences (Leipzig, Germany). All participants reported normal or corrected-to-normal vision, and none of them reported a history of hearing or neurological disorders. Right-handedness was assessed by the Edinburgh Handedness Inventory (Oldfield, 1971).

All participants gave written informed consent and were paid for their participation. The experiment was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the University of Leipzig.

An experimenter (L.V., 26 years, female) was the partner in the social condition. This decision was taken to ensure a natural asymmetrical learning situation in which one partner had knowledge about the game structure. Further, the reason for using the same person to interact with all participants allowed us to control for interaction variability. Although it is possible that the experimenter adapted to each participant unwittingly, it is safe to expect a degree of stability in the experimenter's behavior. To further control for differences in interaction variability (social partner vs. computer), we used measures of variability (standard deviations of response times) as covariates in the statistical analyses of the data.

4.3.2.2 Material

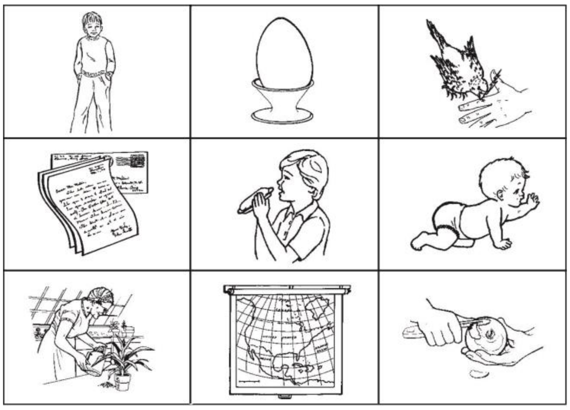
4.3.2.2.1 Checkerboards (pictures)

The stimuli consisted of 240 checkerboards, each containing 9 images (330 width x245 height pixels, 72 dpi) each centered in a different cell of the checkerboard (**Figure 8**). The images depicted black-and-white drawings of objects, humans, animals, or actions selected from a validated database of pictures available on-line (Bates et al., 2003; Szekely et al., 2003, 2004, 2005). A total of 79 images were selected, including 15 pictures representing humans or animals (category: Subject), 24 representing actions (category: Verb), and 40 representing objects, humans, or animals (category: Object). All images represented single objects, humans, or animals. In each checkerboard, two nouns and an action were combined to form a simple transitive German sentence (Noun - Transitive Verb - Target Object. Example: “Der Junge isst das Ei”, “The boy eats the egg”). We define “sentence context” as the combination of Subject and Verb constituents. The pictures representing the elements of a sentence were arranged in such a way that the cells touched each other at least corner to corner. This constrained the game in the following way: (i) only one sentence could be created within each checkerboard, and (ii) only one object could be chosen to form a plausible German sentence based on the sentence context. The other six pictures in each checkerboard (excluding those belonging to the sentence) were distractor images chosen from the initial image pool and were balanced between pictures representing nouns (either animals, humans, objects) and actions. These distractors were selected to ensure that none of them could be considered as an additional plausible object for a given Sentence Context. The checkerboards were further balanced in terms of the mean naming frequency of the items depicted by the 9 pictures; moreover, each element of the target sentence (subject, verb, object) appeared a comparable number of

Figure 8 - Example of checkerboard used in the experiment.

The hidden sentence is in this example composed of the pictures representing a young boy, the act of eating, and the object egg. The sentence “The boy eats the egg” is the only plausible sentence that can be constructed within the given constraints. Elements depicted in the checkerboard are in the first row (top) from left to right: boy (noun), egg (noun), to peck (verb); second row (middle) from left to right: letter (noun), to eat (verb), baby (noun). Third row (bottom) from left to right: to water (verb), map (noun), to peel (verb).

Images are reproduced from http://crl.ucsd.edu/experiments/ipnp/index.html



times in each cell. All possible dispositions of the three target images were employed a comparable number of times.

The stimuli were validated in two studies. In the first study, 20 German native speakers (9F, mean age 24.50 years, SD 2.74) were presented with 582 written transitive sentences in the form subject – transitive verb – object. Participants’ task was to evaluate the plausibility of the sentences and how well the object fitted the sentence context; both judgments were based on a Likert scale from 1 (not plausible/doesn't fit) to 5 (very plausible/fits perfectly). In the second study, the written sentences evaluated in the first study were presented in pictorial form; participants were presented with 3 pictures for each sentence representing subject, verb, and object of the sentence itself. Twenty German native speakers (11F, mean age 24.10, SD 2.94) were asked to a) write the sentence they thought was represented by the picture sequence by typing the sentence into a computer, b) indicate whether such sentence was plausible, and c) indicate whether the image sequence represented the sentence in a good way. Judgments b) and c) were given on the same five-point scale as in the previous experiment. Only sentences and objects with a mean value above 3 in both experiments were included in the final experimental sample. For each image we further calculated the percentage of agreement between the intended meaning of the image and the answer given by the participants in the second validation study (for example, the number of times "dog" was the answer for an image representing a dog). Images with an agreement below 50% were not further considered, unless the choice of the verbal label was a close synonym (e.g., in German, "Arzt" and "Doktor", corresponding to the English "Doctor"; see paragraph 4.1 and 4.2).

4.3.2.2.2 Pseudo-words

A total of 40 objects were chosen as targets for the Sentence Contexts. Each target was associated with a different Italian pseudo-word (length range: min. 4 max. 6 letters) extracted from a set of disyllabic pseudo-words (Kotz et al., 2010). The selected sample of pseudo-words was balanced for syllabic complexity, initial letter, and final letter ("a" or "o"). We excluded words ending in "e" or "i" to avoid possible confounding factors with the Italian plural form, since all the pictures contained singular elements. Each target object and the associated pseudo-word could be presented a maximum of 11 times.

4.3.2.2.3 Experimental conditions

We manipulated 2 factors: 2 levels of social interaction (Social Interaction (S+) and Non-Social Interaction (S-)) and 2 levels of Sentence Context (different sentence Context (dSC) and same Sentence Context (sSC)).

To evaluate the impact of a social partner on the learning process, participants were randomly assigned to one of two conditions: participants assigned to the social condition (S+) performed the task together with the experimenter; participants in the non-social condition (S-) performed the task alone on a computer (see “Task and Experimental Procedure”).

To evaluate the effect of sentence context variability, we split the pool of target objects into two groups: half of the objects (N = 20) occurred repetitively within the same sentence context (sSC – Same Sentence Context). For example, the image representing “cow” was always the correct ending for the same sentence context “the wolf bites”. The other half of the objects (N = 20) was presented at each repetition within a different sentence context (dSC – Different Sentence Context). For example, the image representing “egg” could follow in sentence contexts such as “the woman cuts”, “the boy eats”, etc. Although each sentence was repeated 11 times, the actual number of exposures to each pseudo-word was dependent on the number of correct responses given by the participants, as a pseudo-word was presented only in the case of a correct response (see “Task and Experimental Procedure”).

The two factors were both evaluated as between-subjects factors; thus, every participant was randomly assigned to one of four conditions: Social Interaction and different Sentence Context (S+, dSC - N = 17, 9F, mean age 25.71 years, SD 3.12), Social Interaction and same Sentence Context (S+, sSC – N = 17, 8F, mean age 25.53 years, SD 2.81), no Social Interaction and different Sentence Context (S-, dSC – N = 17, 9F, mean age 24.06 years, SD 3.29), and lastly no Social Interaction and same Sentence Context (S-, sSC – N = 17, 8F, mean age 25.47 years, SD 2.13). There was no age difference between participants in the four groups [$F(3,64) = 1.197$, $p = .318$, $\eta_p^2 = .053$]. In total, each participant was exposed to 20 pseudo-words repeated a maximum of 11 times.

4.3.2.3 Task & Experimental Procedure

The experiment consisted of three parts. First, participants were presented with detailed written instructions of the experiment and performed a block of 10 practice trials to familiarize themselves with the task requirements. Second, after completing the training, participants performed a learning phase. Third, at the end of the learning phase, a testing phase took place to evaluate whether pseudo-words presented during the learning phase had been mapped to the corresponding objects. Participants were not told in advance about this phase. Stimuli were presented using a desktop computer running Presentation 16.0 (Neurobehavioral Systems, Albany, USA). Two standard wheel mice (Logitech Premium Optical Wheel Mouse) were connected to the same Windows computer and used as response devices. Subsequent statistical analyses were performed using MATLAB R2013a (The Mathworks Inc., Natick, USA) and IBM SPSS Statistics 18 (IBM Corporation, New York, USA). The task specifics are described below and displayed in **Figure 9**.

4.3.2.3.1 Practice Trials and Learning phase

In this phase, the participant's task was to find amongst the images of the checkerboard the correct ending for a given context sentence. In all conditions, the trial began with the presentation of a fixation cross (500 ms), followed by a checkerboard. In each checkerboard the participant was provided with the sentence context (**Figure 9**): In the S+ condition, the experimenter selected the Subject and Verb of the sentence by left clicking on them with the mouse in rapid succession. In the S- condition, Subject and Verb were automatically selected by the computer, with a delay of 1,000 ms between the two events. In both conditions, a light blue frame appeared around the picture immediately after its selection and remained on the screen until the participant had provided an answer. In both conditions, the selection of the Verb represented a go signal for the participant to identify the correct target object by left clicking on it with the mouse. There was no time limit. When the answer was correct, the selected image was substituted by a pseudo-word providing the “Italian name” of the object. The pseudo-words were presented in black capital letters on a white background in the cell selected by the participant, and remained on the screen for 1,000 ms (font Arial, size 40 points). Participants were not asked to pay attention to the words or to memorize them. In case of an incorrect response, no “Italian name” was displayed and the following trial began immediately (**Figure 9**).

4.3.2.3.2 Testing phase

The testing phase (**Figure 10**) consisted of two tasks widely employed in learning and memory research: a recognition task and a recall task (Glenberg, 1976, 1979; Lohnas et al., 2011; Polyn et al., 2009; Swanborn & De Glopper, 1999; Verkoeijen et al., 2004). In the recognition task, an object image was presented together with a pseudo-word that participants had seen during the learning phase; participants were asked to indicate whether the association of the two elements was correct or incorrect based on what they had learned during the learning phase. Picture-word associations were correct in 70% of the trials and incorrect in the remaining 30% of the trials. In the recall task, participants were presented with one of the target objects and asked to type in the pseudo-word assigned to the object during the learning phase. No time limit was imposed. The two tasks were presented in counterbalanced order across participants. All participants underwent the same testing phase individually (in other words, without the experimenter), irrespectively of the condition they were exposed to during the learning phase.

Figure 9 – Experimental Trial.

Example of an experimental trial in the social and non-social condition. Abbreviations: SC = Sentence Context.

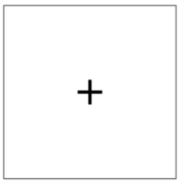
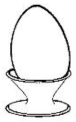
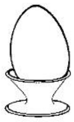
EVENT	SOCIAL Interaction S+	NON social Interaction S-
	Fixation cross Time: 500 msecs	
	The checkerboard appears on computer screen	
	The experimenter selects the SC subject Time: variable	The computer selects the SC subject Time: 1000 msecs
	The experimenter selects the SC verb Time: variable	The computer selects the SC verb Time: 1000 msecs
	The participant selects the target SC object	
	Correct response: the "Italian name" of the object is presented on screen for 1000 msecs Wrong response: next trial	
End Trial		

Figure 10 – Testing phase example.

The picture represents an example of stimuli presented during the testing phase, involving a recognition task and a recall task.

TESTING PHASE	EVENT	TASK
Recognition	 TASNA	Is the pseudo-word the correct "Italian name" of the represented object?
Recall		Type in the "Italian name" of the represented object.

4.3.2.4 Data analyses

Behavioral data were first corrected for outliers; trials with response times exceeding the mean response times (RTs) ± 2 standard deviations (SDs) were excluded from further analysis (mean of rejected trials across participants = 6.33%).

For the learning phase, response times were calculated as the time between the appearance of the "verb" image and the participant's answer. Accuracy scores (proportion of correct responses), response times of correct responses, and their standard deviations were calculated at each repetition of the object for each participant. To evaluate the degree of temporal coordination displayed by the participant during the learning phase, we used the following measures. First, SDs of response times were employed as an index of stability in the participants' performance; in other words, the higher the SDs, the less stable (or more variable) the performance. Further, we calculated the lag-0 and lag-1 cross correlation coefficients between the inter-trial intervals produced by the participants and those produced by the experimenter (S+ condition) or computer (S- condition). These measures are informative about the relation between the temporal behavior of the experimenter/computer and the participant. More specifically, the cross correlation at lag-0 indicates how much the behavior of the participant in one trial is temporally related to the behavior of the partner (experimenter/computer) in the same trial; the cross-correlations at lag-1 indicate whether the behavior of experiment/computer was related to the participant behavior in the following trial.

To account for the difference in the variability of trial presentation in the different conditions, we conducted separate ANCOVAs on the variables of interest using the SDs of experimenter's response/computer times as covariate during the learning phase. We did not covary for SDs in the cross-correlation analyses, since SDs account for the variability in the computer/experimenter RTs series on which the correlation coefficients are calculated.

For the recognition task, response times were calculated as the time between the appearance of the word/image combination and the participant's response; accuracy scores were defined as the proportion of correct responses. For the Recall task, response times of correct responses were calculated as the time between the appearance of the image and the button press to move on to the next trial. Words perfectly recalled (i.e., the recalled pseudo-word was identical to the one presented in the learning phase) were considered correct and assigned 1 point; all other answers were scored as 0. This way of scoring the test is more precise as it tackles only the pseudo-words that are perfectly recalled. However, for completeness, we also calculated the proportion of letters correctly recalled for each recalled/target combination ("l_common").

In both testing tasks, we used the number of exposures during the learning phase as a covariate. This number is directly related to the number of correct responses and takes into account the mean number of times pictures were repeated during the learning phase, ranging from a minimum of 0 (no correct responses) to a maximum of 11 times (no errors).

When the assumption of sphericity was not met, the Greenhouse-Geisser correction was applied to the degrees of freedom. Two-tailed t-tests and simple effect analyses were employed to compare individual experimental conditions. We used an alpha level of $p < .05$ to ascertain significance for all statistical tests.

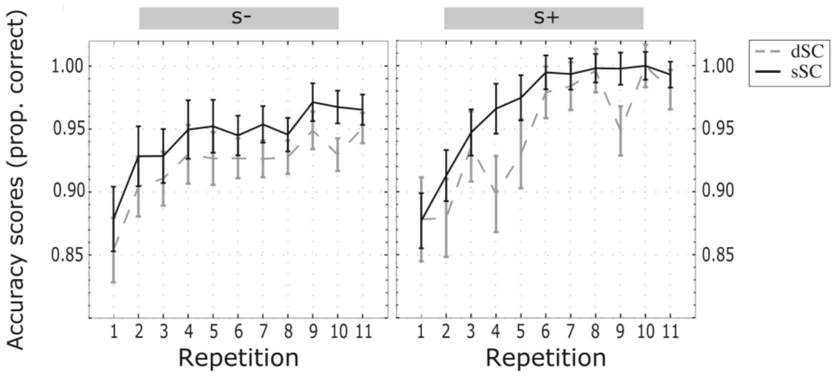
4.3.3. Results

4.3.3.1 Learning phase

Participants achieved an average accuracy of 93%. A 2x2 repeated measures

Figure 11 – Learning phase accuracy scores.

Accuracy of responses in the learning phase plotted as a function of item repetitions while controlling for experimenter/computer variability. Vertical lines represent standard errors of the mean. Abbreviations: S- = non-social interaction; S+ = social interaction; dSC = different Sentence Context; sSC = same Sentence Context.

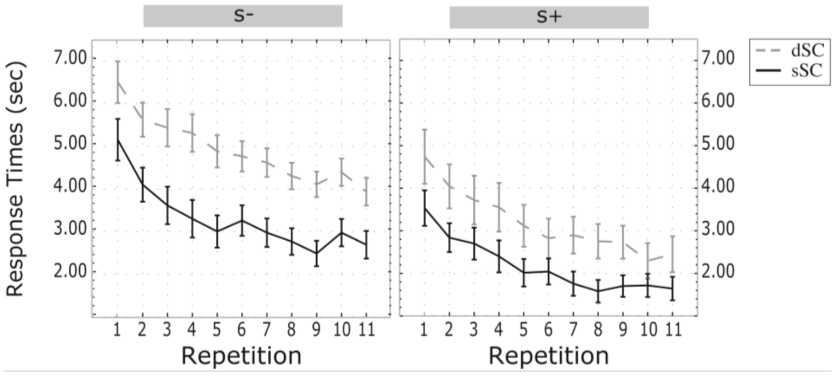


ANCOVA was conducted on the accuracy scores with the between factors Sentence Context (sSC vs. dSC) and Social Context (S+ vs. S-) and the within factor Repetition (11 repetitions). Standard Deviations of presentation times (experimenter, computer) were employed as a covariate to account for the different variability.

Accuracy scores increased over time [main effect of Repetition, linear trend, F(5.78, 364.35) = 13.06, p = .000, η_p^2 = .172] similarly for all experimental conditions (i.e., no interactions, all ps > .102). We observed a marginally significant difference between sSC and dSC [F(1,63) = 3.390, p = .070, η_p^2 = .051], with greater accuracy when words were repeated in same Sentence Contexts [M = .956, SD = .048] compared to different Sentence Context [M = .934, SD = .046]. There was no significant effect of social interaction [F(1,63) = 2.075, p = .155, η_p^2 = .032] and no interaction between the factors [F(1,63) = .000, p = .984, η_p^2 = .000] (**Figure 11**).

Figure 12 – Learning phase response times.

RTs of correct responses plotted as a function of item repetitions while controlling for experimenter/computer variability. Abbreviations: S- = non-social interaction; S+ = social interaction; dSC = different Sentence Context; sSC = same Sentence Context.

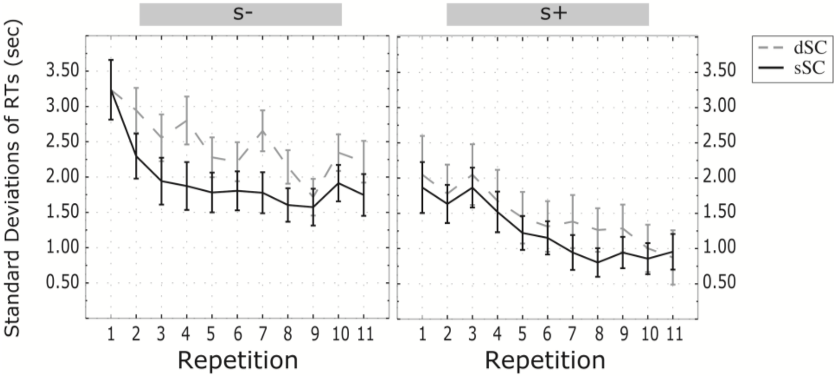


RTs of correct responses significantly decreased over the course of the learning phase [main effect of Repetition, linear trend, F(4.02, 253.31) = 31.538, p = .000, η_p^2 = .334] similarly for all experimental conditions (no interactions, all ps > .521). Furthermore, participants were faster when learning socially (S+, M = 3.013, SD = 1.310) as compared to non-socially (S-, M = 3.711, SD = 1.486) [F(1,63) = 9.496, p = .003, η_p^2 = .131]. For Sentence Context, RTs were faster in the sSC (M = 2.567, SD = 1.165) as compared to the dSC (M = 4.157, SD = 1.232) condition [F(1,63) = 17.345, p = .000, η_p^2 = .216]. There was no interaction between the factors [F(1,63) = .846, p = .361, η_p^2 = .013] (**Figure 12**).

Participants' ***variability*** (as expressed by the ***standard deviations of response times***) decreased over time [linear trend, F(6.30, 396.56) = 7.610, p = .000, $\eta_p{}^2$ = .108] in a comparable way for all experimental conditions (no interactions, all p >.377). Participants learning in a social context (S+, M =

Figure 13 – Learning phase standard deviations.

SDs of response times indicating participants' variability plotted as a function of item repetitions while controlling for experimenter/computer variability. Vertical lines represent standard errors of the means. Abbreviations: S- = non- social interaction; S+ = social interaction; dSC = different Sentence Context; sSC = same Sentence Context

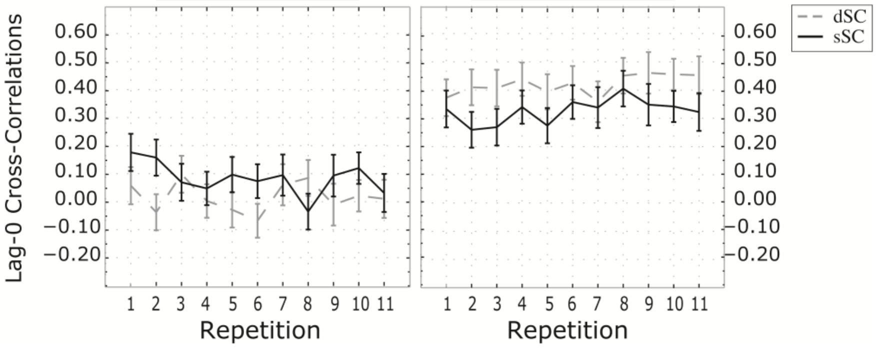


1.560, SD = .738) were less variable than participants learning non-socially (S-, M = .201, SD = .958) [F(1,63) = 7.434, p = .008, η_p^2 = .106]. As for the Sentence Context, there was no difference in this analysis between sSC and dSC [F(1,63) = 2.585, p = .113, η_p^2 = .039]. The interaction between these two factors was not significant [F(1,63) = .024, p = .424, η_p^2 = .007] (**Figure 13**).

We further evaluated the degree of temporal correlation between the two inter-trial-interval (ITI) time series (experimenter/computer – participant). The ANOVA conducted on the ***lag-0 cross-correlations coefficients*** shows that this index did not change over time (no main effect of Repetition, F(10, 640) = .309, p = .979, η_p^2=.005). Participants learning in the social condition (S+, M = .301, SD = .028) had higher correlation coefficients compared to S- participants (M = .129, SD = .028) [F(1,64) = 97.212, p = .000, η_p^2 = .603]. There was no significant difference between sSC and dSC [F(1,64) = .200, p = .656, η_p^2 = .003]. There was, however, a significant interaction between Sentence

Figure 14 – Learning Phase lag-0 cross-correlations coefficients.

Lag-0 ccs are plotted as a function of item repetitions separately for non-social and social conditions (respectively: left and right panel). Vertical lines represent standard errors of the means. Abbreviations: S- = non-social interaction; S+ = social interaction; dSC = different Sentence Context; sSC = same Sentence Context



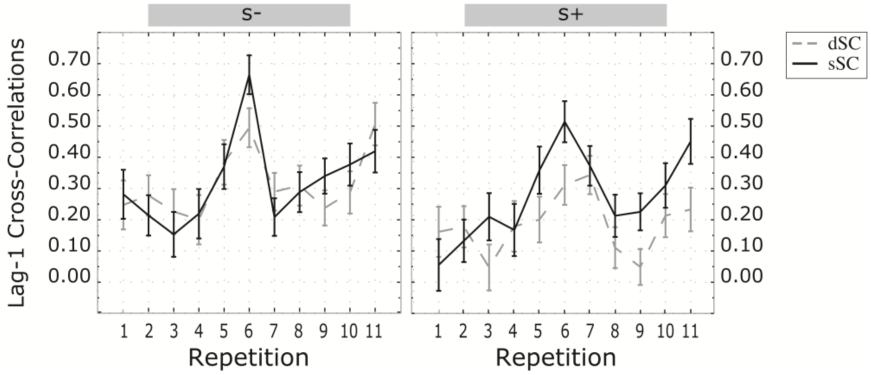
Context and Social Context [F(1,64) = 6.178, p = .016, $\eta_p{}^2 = .088$]. Thus, we computed a simple effect analysis showing that there was no difference between the levels of sentence context in the non-social group [F(1,64) = 2.078, p = .154]. However, in the social group, the correlation coefficients were higher in the different Sentence Context condition [M = .425, SD = .033] than in the same Sentence Context condition [M = .329, SD = .031] [F(1,64) = 4.300, p = .042] (**Figure 14**).

The analysis of the ***cross-correlations at lag-1*** showed a main effect of Repetition [polynomial of 4th order, F(10,640) = 10.015, p = .000, $\eta_p{}^2 = .137$). The S+ condition had smaller correlation coefficients at lag-1 [M = .229, SD = .130] as compared to the S- [M = .317, SD = .117] conditions [F(1,64) = 3.714, p = .004, $\eta_p{}^2 = .124$]. There was no effect of sentence context and there were no interactions (all ps > .108) (**Figure 15**).

To summarize, while accuracy levels were comparable across conditions,

Figure 15 – Learning phase lag-1 cross-correlations coefficients.

Lag-1 ccs are plotted as a function of item repetitions separately for non-social and social conditions (respectively: left and right panel). Vertical lines represent standard errors of the means. Abbreviations: S- = non-social interaction; S+ = social interaction; dSC = different Sentence Context; sSC = same Sentence Context.



significant differences emerged in the temporal behavior of participants as a function of the type of interaction (social or non-social) to which they were exposed: Standard deviations, response times, and cross-correlations of ITIs all showed significant differences between participants learning socially and non-socially. Sentence context variability influenced response times, with lower RTs for words embedded in a consistent sentence context. Further, only for the participants learning socially, words embedded in a more variable sentence context yielded higher values of lag-0 cross-correlations.

4.3.3.2 Testing phase

4.3.3.2.1 Recognition task

In general, participants performed well in the recognition task. The mean accuracy of 87% was significantly above chance level [1 sample t-test against a chance level of .50, t(67) = 22.772, p = .000, 95% C.I. (.333, .397); the chance level was set at .50 since the task required a yes/no answer].

A separate 2 (Social Context: S+ vs S-) x 2 (Sentence Context: sSC vs dSC) ANCOVA was conducted on Accuracy scores and Reaction Times to evaluate the impact of the experimental manipulations while accounting for the number of exposures to the pseudo-word during the learning phase.

Accuracy scores did not differ between participants learning socially and participants learning on a computer [F(1,63) = 1.005, p = .320, η_p^2 = .016], nor between sSC and dSC [F(1,63) = .690, p = .409, η_p^2 = .011]. However, the interaction between the factors was significant [F(1,63) = 5.965, p = .017, η_p^2 = .086]. Therefore, an analysis of simple effects was carried out. This analysis revealed that sSC words were recognized significantly better by participants learning non-socially [M = .901, SD = .304] as compared to participants learning socially [M = .793, SD = .034] [F(1,63) = 6.111, p = .016]. Rather, dSC words were similarly recognized by both exposure groups [F(1,63) =

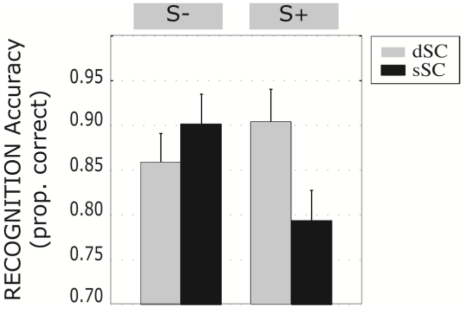
1.022, p = .316]. Participants who learned socially recognized dSC words better [M = .904, SD = .036] than words repeated in the same sentence context [M = .793, SD = .034] [F(1,63) = 4.114, p = .047], an effect not found in the non-social group [F(1,63) = .761, p = .386] (**Figure 16**).

The ANCOVA conducted on the ***RTs*** did not show any significant effect for Social Context [F(1,63) = 0.57, p = .812, $\eta_p{}^2 = .001$] or Sentence Context [F(1,63) = .029, p = .866, $\eta_p{}^2 = .000$], and no interaction of both [F(1,63) = .753, p = .389, $\eta_p{}^2 = .012$].

To evaluate whether the number of words recognized was related to the degree of temporal coupling during the learning phase, we ran a partial correlation analysis between accuracy scores collected in the Recognition task and the correlation coefficients at lag-0 and lag-1, still controlling for the mean number of repetitions. Both analyses turned out to be non-significant (all ps > .431).

Figure 16 – Recognition phase accuracy scores.

Vertical lines represent standard errors of the mean. Abbreviations: S- = non-social interaction; S+ = social interaction; dSC = different Sentence Context; sSC = same Sentence Context.



In summary, recognition accuracy scores revealed a significant interaction between sentence context and social interaction: participants who learned non-socially recognized words encoded in a consistent context better, while participants who learned socially recognized words encoded in a variable context better.

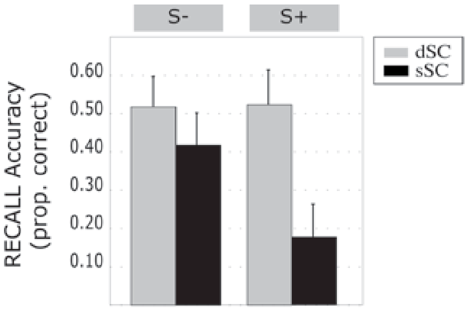
4.3.3.2.2 Recall task

The recall task was more challenging for the participants as revealed by the overall low number of correctly recalled items (40%). When using a more liberal criterion of the number of letters correctly recalled for each pseudo-word ("l_common"), the percentage of correctly recalled items increased to 60%. In this analysis we did not define a chance level as this task required a complex answer (that is, typing in the correct pseudo-words on a keyboard).

A Separate 2 (Social Context: S+ vs S-) x 2 (Sentence Context: sSC vs dSC) ANCOVA was conducted on ***Accuracy scores*** while accounting for the mean

Figure 17 – Recall phase accuracy scores.

Vertical lines represent standard errors of the mean. Abbreviations: S- = non-social interaction; S+ = social interaction; dSC = different Sentence Context; sSC = same Sentence Context.



number of repetitions participants were exposed to during the learning phase. This analysis yielded a significant difference for Sentence Context, with dSC revealing higher scores [M = .521, SD = .064] than sSC [M = .298, SD = .067] [F(1,63) = 4.671, p = .034, ηp^2 = .069]. However, there was no effect of Social Context [F(1,63) = 2.232, p = .140, ηp^2 = .034], nor an interaction between the two factors [F(1,63) = 2.456, p= .122, ηp^2 = .038] (**Figure 17**). There were no significant results when repeating this analysis on the *l_common* index (all ps > .064).

There were no significant differences for ***RTs*** in the recall task for any of the experimental conditions, nor the interactions between the factors (all p > .311).

To evaluate whether the number of words recalled was related to the degree of temporal coordination during the learning phase, we ran a partial correlation analysis between accuracy scores during the Recall task and the correlation coefficients at lag-0 and lag-1 while controlling for the mean number of repetitions. Both the analyses turned out to be non-significant (all ps > .400)[16].

In summary, in the recall task, more words encoded in a variable context were correctly recalled than words encoded in a consistent context.

4.3.4. Discussion

The aim of the current study was to investigate the impact of social interaction on adult contextual word learning. We addressed this question by implementing an ecologically plausible set-up based on contextual learning principles. In this task, participants learned new words either alone or with a knowledgeable partner; these new words were embedded in sentence contexts with different degrees of variability. Our results show that during the learning

---

[16] Participants in the social condition always performed the task with the same female experimenter. Although the presence of female and male participants was fully counterbalanced across all conditions, we checked whether there were gender differences in the dependent variables. We did not find any difference between male and female participants for any of the dependent variables (all ps > .150).

phase social interaction significantly influences participants' temporal coordination: standard deviations of response times were smaller, response times were faster, and temporal correlations higher when participants were learning socially compared to non-socially. Furthermore, in the social group, temporal correlations with the partner were higher when words were presented in variable sentence contexts. Variable contexts further elicited slower response times compared to a consistent context. In a subsequent recognition task, participants who learned socially recognized more words when they were originally embedded in a variable context. Participants who learned with a computer recognized more words embedded in a same context compared to participants who learned with the experimenter. Lastly, in a recall task, more words embedded in different contexts during learning were recalled than words repeated in the same context.

Learning a new language is a very complex task; however, most often people are able to learn words in a new language relatively easily and effortlessly. How is this possible? The literature on first language acquisition emphasizes the role of the caregiver, who seems to direct a child's attention to the correct referents of new verbal labels (Csibra & Gergely, 2009; Frith & Frith, 2006; Louwerse et al., 2012; Rader & Zukow-Goldring, 2012). In this process, temporal coordination between the child and the caregiver is crucial, as the learner's attention needs to be guided to the correct referent at the correct point in time (Pereira et al., 2008; Rolf et al., 2009). Adults learning new words in a social context face a problem remarkably similar to children: when a new word is encountered, it is necessary to understand what the word means. In other words, a referent for the new verbal label needs to be specified. We hypothesized that the temporal coordination with a partner during social interaction (for example, Richardson, Marsh, Isenhower, Goodman, & Schmidt, 2007) facilitates the learning process in adults in a similar fashion to the learning process in children. In the following, we will discuss the current results in three steps.

First, we will discuss the results, which revealed that temporal coordination of learning partners emerges in contextual learning. Second, we will focus on whether this form of temporal coordination facilitates word learning. Third, we will consider how other sources of information (for example, variable sentence context) contribute to this learning process.

The results of the learning phase constitute the first step. In line with our hypotheses, temporal coordination was higher during social learning than during non-social learning. This result extends previous evidence reporting that partners involved in social interactions tend to coordinate their behavior to achieve common goals (Hasson et al., 2012; Schmidt et al., 2011; Stivers et al., 2009; M. Wilson & Wilson, 2005; Yun et al., 2012). Such temporal coordination is likely supported by the human ability to infer a social partner’s intention during an interaction (Frith & Frith, 2006, 2012). Indeed, understanding a partner's intention is essential to predict what will happen next (Frith & Frith, 2012; Verga & Kotz, 2013) and, consequently, to adjust the time-course of one’s own action (Pecenka et al., 2013; Vesper, van der Wel, Knoblich, & Sebanz, 2011). However, human interactions entail much more than simply predicting a next event and adjusting to it. If this were the case, temporal coordination would have been maximal with a perfectly predictable partner. Participants performing the computer-based learning phase faced this situation; yet, despite interacting with a highly predictable “partner” (the computer), participants in this learning condition showed significantly lower temporal coordination than participants dealing with a less predictable human social partner. Indeed, participants socially interacting not only displayed better temporal coordination with the experimenter but also were more precise in their behavior as indicated by smaller standard deviations in their response times. Reduced variability in the participants' performance is a typical observation in joint action studies and is interpreted as an indication of the fine-tuning of one’s own action necessary to ensure smooth coordination between partners (Repp, 2005; Vesper et al., 2011, 2012; Yun et al., 2012).

Importantly, in the current study the lag-1 cross-correlations between the participant and the experimenter/computer (i.e., the similarity between the behavior of the participant and what had occurred in the previous trial) was lower in participants learning socially compared to participants learning non-socially on a computer. This evidence further corroborates the interpretation of an on-line temporal coordination between participant and experimenter: indeed, temporal coordination emerges as a mutual adaptation between two persons interacting together to achieve a common goal (in this case, to identify the sentence represented in the checkerboard).

In addition to these results, we also observed that participants were faster when interacting with an experimenter than with the computer. This result may suggest that temporal coordination facilitates the guiding of attention of the participant to the correct referent of a new word; consequently, the participant needs less time to identify the target referent since attention is already allocated to it. This interpretation parallels findings in first language acquisition studies, suggesting that improved temporal coordination between a learner and a more knowledgeable person directs the learner's attention towards correct new word referents (Pereira et al., 2008). However, another possible explanation is that faster reaction times reflect a strategy adopted by the participant to reduce the gap between the experimenter’s and her/his own activity. This interpretation fits with the general rules underlying turn-taking, which are intended to minimize the distance between turns while, at the same time, avoiding overlaps (Stivers et al., 2009). Further studies are required to disentangle these possible interpretations. Taken together, the results of the learning phase support the hypothesis that temporal coordination increases during social interactions but not in non-social interaction during contextual word learning.

Second, we addressed the question whether increased temporal coordination per se facilitates word learning. The results of the testing phase do not confirm this hypothesis. In both tasks employed to test new word-referent mapping, participants who trained socially did not perform better than participants who learned non-socially. However, some task-specific aspects with regards to the learning outcome in the testing phase need to be further considered.

First, the high accuracy scores obtained in the recognition task suggest that pseudo-word – referent association was successful in most of the newly learned associations. Thus, it may be possible that the task was too easy to uncover differences dependent on the type of learning phase to which participants were exposed. By employing a test of intermediate difficulty between recognition and recall, we may be able to answer this question. Some contextual learning studies used such an intermediate test in which participants were exposed to plausible or non-plausible sentences containing a target word (for example, Borovsky et al., 2010). This type of test would also be more in line with the current learning phase in which complete sentences were presented.

Moreover, an additional disadvantage between the learning and the testing phase may have been the very type of interaction in which participants were learning new words. While learning new words in a computer interface and testing participants in a computer interface is consistent, participants who learned in social interaction may have been disadvantaged as they experienced a contextual inconsistency between the learning and the testing phase. For example, consistency between learning and testing environments has been suggested to facilitate recall (Godden & Baddeley, 1975; Polyn, Norman, & Kahana, 2009). Accordingly, participants engaged in a non-social learning environment had an advantage in the testing phase, as recall during testing was similar to learning. This explanation may shed light on the differences between second and first language learning. As stated in the introduction, word learning is facilitated in children when there is good temporal coordination with the caregiver during a word learning game (Pereira et al., 2008). However, in first language studies the learning and testing phases are always contextually consistent, since an adult (either the caregiver or the experimenter) is with the child during both phases. For this reason, it cannot be concluded that social word learning in adults is different from first language learning. Further studies are required to elucidate whether contextual inconsistencies between learning and testing are the cause of this null result. Thus, while the current study aimed to provide first evidence on adult social and contextual word learning, additional evidence is needed to reveal the nature of the mechanisms underlying this type of learning.

Crucially, the third step in this discussion evaluates the combined effects of social interaction and variable sentence context. In the current study, we exploited contextual learning in social interaction to re-create a natural learning situation. Using a well-known learning set-up, this interaction allowed us to investigate whether social and contextual cues are integrated during adult word learning.

During the learning phase, words repeated in a consistent sentence context elicited faster response times compared to words embedded in more variable sentence contexts. This result can be explained as a contextual cueing effect: attention is guided by repetitive contextual cues; as a consequence, repeated patterns facilitate the orienting of attention to the correct target (Chun & Jiang, 1998; Conci & Müller, 2012). Nevertheless, words embedded in variable sentence contexts led to higher recall scores in the testing phase. We hypothesized that when words appear in variable contexts, each embedding of a target word leads to more available cues that enrich the representation of the new word’s meaning (Adelman et al., 2006; Verkoeijen et al., 2004). Evidence on increased recall rates for words originally embedded in different contexts over words repeated in a consistent context is therefore in line with our hypothesis.

While these results are an important confirmation of previously reported effects of word learning, the current results also critically extend this evidence by providing first data on whether and how cues derived from social interaction and sentence context are utilized by the learner.

Our results suggest that social and contextual cues are not used independently, but are merged to direct the learner's attention to the target referents of a new word.

In the social group, but not in the non-social group, temporal coordination between the participant and the experimenter was higher when new words appeared in different sentence contexts and lower when words were embedded and repetitively presented in the same context. A possible interpretation for this result is that if a new word is repeated in different contexts, it is not possible to apply a priori knowledge of the correct word referent. Consequently, the learner may have to rely on the partner to obtain cues as to which referent may be the correct one. This situation is somewhat similar to that of a child learning the first words: the learner is faced with a constantly changing environment (in our case, the checkerboard), in which multiple referents are present. In such a situation, the help of the caregiver is critical for children in their effort to learn new words (Csibra & Gergely, 2009; Kuhl, 2007) and is supported by the temporal coordination emerging between partners of a dyad (Pereira et al., 2008). Similarly, adult learners may rely more on a social partner when they have to find a referent for a new word in a given learning environment. On the other hand, words repeated in the same sentence context do not require the experimenter to guide the learner's attention towards the target as the target will be known from previous presentations in the same sentence context. Crucially, as we predicted, in the social learner group words originally encoded in a variable sentence context were also recognized better than words encoded in a repeated context. This result is particularly interesting in light of the hypothesis that smooth coordination creates a sort of "social rhythm" (Pereira et al., 2008), which may represent a special form of a multimodal rhythm. Multimodal rhythms emerge when movements are coordinated with an external visual or auditory stimulus. In the case of social learning, the role of the external stimulus may take the form of the learner's social partner. It has been shown that such multimodal rhythms facilitate the allocation of attention and the integration of information pivotal to successfully learn new information (Lagarde & Kelso, 2006; Rolf et al., 2009; Schmidt-Kassow et al., 2013). The lack of a significant correlation between temporal coordination and testing scores is not conclusive in this regard as it may be the case that the testing phase employed in this study was not sensitive enough to capture learning effects.

To conclude, the current study aimed to investigate the impact of social interaction on word learning in young adults. By exploiting a contextual learning set-up, we were able to show that adults use cues derived from social interaction to coordinate their behavior with a partner. Further, such coordination interacts with cues derived from the context the new word is presented in to determine the learning outcome. While providing first evidence on how social interaction may influence word learning, this study opens a major avenue for future research in this domain. For example, word learning in children seems to require the presence of a real partner; a video-recording of a caregiver does not have the same impact as a real person (Kuhl et al., 2003). A relevant question for future research is whether the physical presence of another person is necessary in adult learners as well. More specifically, further studies are required to elucidate what exactly drives temporal coordination with a social partner (for example, visual or auditory feedback).

4.4 Play along: Effects of music and social interaction on word learning[17]

Abstract

Learning new words is an increasingly common necessity in everyday life. External factors, among which music and social interaction are particularly debated, are claimed to facilitate this task. Due to their influence on the learner’s temporal behavior, these stimuli are able to drive the learner's attention to the correct referent of new words at the correct point in time. But do music and social interaction impact learning behavior in the same way? The current study aims to answer this question. Native German speakers (N = 80) were requested to learn new words (pseudo-words) during a contextual learning game. This learning task was performed alone with a computer or with a partner, with or without music. Results showed that music and social interaction had a different impact on the learner’s behavior: Participants tended to temporally coordinate their behavior more with a partner than with music, and in both cases more than with a computer. However, when both music and social interaction were present, this temporal coordination was hindered. These results suggest that while music and social interaction do influence participants’ learning behavior, they have a different impact. Moreover, impaired behavior when both music and a partner are present suggests that different mechanisms are employed to coordinate with the two types of stimuli. Whether one or the other approach is more efficient for word learning, however, is a question still requiring further investigation, as no differences were observed between conditions in a retrieval phase which took place immediately after the learning

---

[17] This study is based on the article: Verga, L., Bigand, E. & Kotz, S.A. (in preparation) Play along: Temporal coordination to music and social interaction in second language vocabulary learning.

session. This study contributes to the literature on word learning in adults by investigating two possible facilitating factors, and has important implications for situations such as music therapy, in which music and social interaction are present at the same time.

4.4.1. Introduction

In an increasingly multicultural world, even adult speakers often face the necessity to acquire a foreign language starting from its building blocks: words. New words are frequently encountered in everyday life, and the first step to learning them is to understand what they mean. However, possible meanings for a new verbal label are countless. How does the learner identify the correct one? Research in second language learning has identified several factors that may facilitate learners in their effort to acquire new vocabulary, among which music and social interaction stand out as particularly important, yet their role is still debated.

The idea that music may boost language functions has fascinated the scientific community for quite some time (Schellenberg, 2003), with particularly convincing evidence coming from clinical studies (Altenmüller & Schlaug, 2013; de l' Etoile, 2010; Hillecke, Nickel, & Bolay, 2005; Hurkmans et al., 2011; Simmons-Stern, Budson, & Ally, 2010; Thaut, 2010; Thompson, Moulin, Hayre, & Jones, 2005). Similarly, in healthy populations several studies report a positive effect of music on the encoding and decoding of verbal material, with music being used either as a background (De Groot, 2006; Ferreri et al., 2014; Ferreri, Aucouturier, Muthalib, Bigand, & Bugaiska, 2013), as a contrast for sung and spoken material (Ludke, Ferreira, & Overy, 2014; Rainey & Larsen, 2002), or as a form of long-term training (Ho, Cheung, & Chan, 2003; Kilgour, Jakobson, & Cuddy, 2000). The question remains open, however, as to which specific aspects of music impact learning. It has been proposed that the boosting effect of music may depend on different mechanisms (for example, temporal scaffolding/attention, emotion/reward and arousal/mood), recruited by progressively higher levels of musical complexity (Ferreri & Verga, in prep.). In particular, this account suggests that simple musical stimuli aligned with verbal material may significantly potentiate learning by providing a temporal structure, in which temporal regularities orient participants' attention to the verbal information to be encoded (Francois & Schön, 2010; Jones & Boltz, 1989; Schön et al., 2008; Thaut, Peterson, & McIntosh, 2005); in the case of vocabulary learning, this information is represented by new words and their respective referents. By facilitating predictions of "what is coming next" (Collins, Tillmann, Barrett, Delbé, & Janata, 2014; Mathias, Palmer, Perrin, & Tillmann, 2014; Tillmann, Janata, & Bharucha, 2003), the temporal regularities conveyed by music also induce temporal coordination[18]. Indeed, a tight link between music and coordinated motor behavior emerges very early on in life (for example see Phillips-Silver & Trainor, 2005) and continues throughout the entire lifespan, as demonstrated by the fact that listeners often "tap their feet or nod along to the beat of a tune" (Chen, Penhune, & Zatorre, 2008; see also Loehr, Large, & Palmer, 2011; Repp & Su, 2013). Importantly, this form of auditory-motor synchronization to music has been shown to further improve attentional processing, by facilitating the temporal encoding of the stimuli (Schmidt-Kassow, Heinemann, Abel, & Kaiser, 2013).

Interestingly, similar mechanisms (that is, attention orienting and temporal coordination) have been proposed to explain the facilitating effect of social interaction on word learning in children, for whom the presence of another person is a *sine qua non* condition to build up new vocabulary (Kuhl, 2007;

---

[18] For consistency, the term "temporal coordination" is used in this paper to describe the establishment of temporal dynamics between participants and music or a social partner. It must be pointed out that the same phenomena are described by other authors as synchronization (for example Yun, Watanabe, & Shimojo, 2012), entrainment (for example Knoblich & Sebanz, 2008), mutual adaptation (for example Konvalinka, Vuust, Roepstorff, & Frith, 2010) or with the more general term coupling (for example Demos, Chaffin, Begosh, Daniels, & Marsh, 2012).

Kuhl, Tsao, & Liu, 2003). In these asymmetric learning settings the role of the more experienced person is to guide the learner's attention towards the correct referent for a new word, thus strongly reducing the number of possible referents (Csibra & Gergely, 2009; Hirotani et al., 2009). For this facilitation to occur, temporal coordination between the learner and the social partner is required to triangulate attention towards the target referent at the correct point in time (Gogate et al., 2000; Rader & Zukow-Goldring, 2012; Rolf, Hanheide, & Rohlfing, 2009). Support for this claim comes from evidence that children learn significantly more new words when they are able to reach a good temporal coordination with their caregiver (Pereira et al., 2008). However, as social verbal learning in adults has not been the focus of research until recently (Jeong et al., 2010, 2011; Verga & Kotz, 2013), the impact of a partner on second language acquisition still remains an open question. Similarly to children, coordination with a more experienced partner may create a sort of "multi-modal rhythm" capable of facilitating the allocation of attention and the binding of information required for learning (that is, the correct referent and its new verbal label; Lagarde & Kelso, 2006; Rolf et al., 2009). While the emergence of spontaneous temporal coordination during interactive social situations is frequently and reliably reported in literature on joint action (for example Richardson, Marsh, Isenhower, Goodman, & Schmidt, 2007; Yun, Watanabe, & Shimojo, 2012), its impact on word learning has not yet been investigated.

The evidence reported so far suggests that common properties in music and social interaction – such as the establishment of a temporal structure – may boost word learning by facilitating the allocation of attention and the emergence of spontaneous temporal coordination. Importantly, however, these are not the only commonalities between music and social interaction that justify a comparison between the two stimuli: Indeed, they are both rich, complex stimuli that are pleasurable and enjoyable (Blood, Zatorre, Bermudez, & Evans, 1999; Hari & Kujala, 2009), which are often concurrently present in a number of contexts (for example musical performance, music therapy). Nevertheless, an important distinction needs to be made: Listening to music has a unidirectional influence, in the sense that the listener coordinates with the music, but not vice-versa (Repp & Keller, 2008). Instead, social interaction elicits a bidirectional influence between partners, who tend to reciprocally modify their behavior (Richardson et al., 2007; Yun et al., 2012). In this scenario, predictions about what is coming next need to be constantly updated in order to allow the adaptation of one's own behavior, an ability critically dependent on the typically human skill to infer the other person's intentions (Frith & Frith, 2006, 2012). Whether this difference influences the way temporal coordination is achieved represented the topic of a recent study by Demos and colleagues (Demos et al., 2012). In their experiment, these authors evaluated participants' coordination with music or a partner, while seated in rocking chairs, and observed that spontaneous coordination emerged with music as well as with a partner. However, coordination with music was weaker than with a partner. Further, when both music and the partner were present, they competed as sources of attraction, resulting in a weaker coordination. The authors interpret these results by proposing that coordination with music differs from coordination with a partner because people interacting together behave as coupled oscillators (Demos et al., 2012; Dumas et al., 2010; M. Wilson & Wilson, 2005). At the neural level, this behavior reflects the activity of populations of neurons in the member of the dyad which become synchronized in their oscillating firing pattern (Cui et al., 2012; Dumas et al., 2011, 2010; Hasson et al., 2012), and in turn, this coupling is reflected in temporal coordination emerging at the behavioral level (Pereira et al., 2008; Richardson et al., 2007; Yun et al., 2012). From a psychological standpoint, these phenomena create a "common ground" between partners, facilitating the transmission of information (Csibra & Gergely, 2009). While this psychological state has been deemed pivotal for children to determine the adult's referent of a new word (Tomasello, 2000), whether adult learners may also benefit from this "shared ground" is still an open question (Jeong et al., 2010; Pickering & Garrod, 2004; Stephens, Silbert, & Hasson, 2010; Verga & Kotz, 2013). On the one hand, the presence of a knowledgeable partner may help to reduce the number of possible referents for a new word; on the other hand, adults do posses – compared to infants – more refined cognitive abilities, which may be sufficient for acquiring new words. In a previous study, we investigated this matter in a social/non-social word learning game (Verga & Kotz, in prep.). Our results suggest an intermediate position: Temporal coordination does emerge between interacting partners, but indeed depends on task demands. If the context, in which a new word is encountered, is particularly challenging (for example, when it is not possible to know a priori the correct referent because the context of a word presentation changes at each occurrence) temporal coordination with a partner is greater compared to less demanding learning situations (for example, when the context is always the same at each occurrence of the new word). As a consequence, participants learning with a partner remembered more words when they were originally presented in different contexts. This result per se does not rule out the possibility that a social partner simply provides a temporal structure able to drive participants' attention towards the verbal information to be encoded; in this case, the source of information should be irrelevant. Conversely, if the establishment of a "common ground" – partially reflected by temporal coordination between the partners – is as important in adult learners as it is in infants, then social interaction should provide an advantage when compared to other forms of temporally structured stimuli, such as music. In other words, this corresponds to the question of whether it is necessary for this temporal structure to be conveyed by *someone,* or if it is enough for it to be conveyed by *something*.

In the current study, our aim was to answer this question by implementing a social/non-social contextual learning task that could be performed either with or without music. Based on the literature reviewed above, we expected participants to achieve better temporal coordination with a social partner (Richardson et al., 2007; Yun et al., 2012) and with music (Demos et al., 2012; Repp & Keller, 2008) when compared to a computer, but hindered when both music and social interaction were present (Demos et al., 2012). Indeed, as suggested above, music and social partners exert different influences (unidirectional versus bidirectional) on participants, possibly implemented by different mechanisms (temporal regularities versus common ground). When both music and a social partner are present, participants either have to integrate the two sets of information or choose just one set and ignore the other. In terms of word learning, if the establishment of a "common ground" is essential, then an improved word-learning rate should be observed in the social interaction condition, regardless of the fact that music also drives the learner's attention toward the correct referent for new words. Instead, if this latter aspect is what drives word learning, then no difference should be observed between music and social interaction. However, it may still be the case that neither music nor social interaction provides useful cues at all, as adult learners are cognitively equipped to learn new words without any additional help. To investigate this hypothesis, we manipulated the variability of the sentence context in which new words were embedded to obtain a "difficult" condition (that is, words were repeated in a different context so the word referent had to be identified ex-novo at each occurrence) and an "easy" condition, in which task requirements were less demanding (that is, words were repeated in the same sentence context so the referent was already known from previous presentations of the same word). In line with our previous results, we expected music and social cues to be maximally used in the "difficult" condition, but not used in the "easy" condition.

4.4.2. Methods

4.4.2.1 Participants

80 native German speakers (40F, mean age 24.86 years, SD 2.62) took part in the experiment. They were all recruited from a database from the Max-Planck Institute for Human Cognitive and Brain Sciences (Leipzig, Germany). All participants reported normal or corrected to normal vision, and none of them reported a history of hearing or neurological disorders. Right-handedness was assessed by the Edinburgh Handedness Inventory (Oldfield, 1971). An experimenter (LV, F, 28 years) was the partner in the social interaction conditions. All participants gave written informed consent and were paid for their participation. The experiment was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the University of Leipzig.

4.4.2.2 Material and apparatus

4.4.2.2.1 Visual stimuli: checkerboards and pseudo-words

Visual stimuli consisted of 180 checkerboards (3 x 3) each containing 9 images (330 x 245 pixels, 72 dpi) each centered in a different cell of the checkerboard (**Figure 8**). The images were black and white drawings representing objects, humans, animals or actions selected from a validated database available online (Bates et al., 2003; Szekely et al., 2003, 2004, 2005; http://crl.ucsd.edu/experiments/ipnp/). A total of 49 images were employed, including 12 pictures representing humans or animals (category: Subject), 17 representing actions (category: Verb), and 20 representing objects, humans or animals (category: Object). All images represented single objects, humans or animals.

In each checkerboard, two nouns and an action were combined to form simple transitive German sentences (noun - transitive verb - target object; for example, “Der Junge isst das Ei”, “The boy eats the egg”). We defined the combination of subject (“Der Junge) and verb (“isst”) as the “sentence context”. Images depicting elements of the sentence were represented in cells touching each other at least corner to corner. Given this constraint, only one object could be chosen to form a plausible German sentence. The six pictures not belonging to the target sentence were distractor images chosen from the initial image pool and were balanced between nouns (either animals, humans or objects) and actions. None of these distractor images constituted a plausible object for the given sentence context. The checkerboards were further balanced for mean naming frequency of the depicted items and mean number of times each element of the target sentence (subject, verb, object) appeared in each cell. All possible dispositions for the three target images were employed a comparable number of times. Details of the stimuli validation have been described elsewhere (Verga & Kotz, in prep.).

Images belonging to the category “objects” (N = 20), which were employed as targets for the sentence context, were each associated with a different pseudo-word. These stimuli were based on Italian word structure and were selected from a published set of disyllabic pseudo-words (Kotz et al., 2010). The selected pseudo-word sample (length range: min. 4, max. 6 letters) was balanced for syllabic complexity, initial letter and final letter (“a” or “o”). We excluded words ending in “e” or “i” to avoid a possible confound with the Italian plural form, since all the pictures contained singular elements. Each pseudo-word and the associated target object could be presented a maximum of 9 times during the learning phase of the experiment.

4.4.2.2.2 Auditory stimuli: melodies

Two novel melodies were created ad hoc to comply with our requirements. One melody was assigned to the "subject" of the sentence context, while the other melody was assigned to the "verb". The melodies needed to parallel the role of the experimenter in the social condition as closely as possible. For this reason, the following criteria were applied: First, the length (duration) of the "subject melody" was adjusted to be comparable to the response times of the experimenter in the social interaction data previously collected. Thus, while the original melody was always the same, we ended up with 9 different tempi, and progressively faster tempi were used from the first to the last repetition. The duration of the musical excerpts ranged from 3.95 to 4.29 seconds. To allow comparisons with the computer condition, the same durations were applied to jitter the stimuli in the silent condition. Second, a melody was created for the "verb" picture with a fixed duration of 600 ms. This duration was comparable to the response times of the experimenter (which were extremely stable over the course of the experiment) to provide the "verb" picture. Third, the choice of a single melody for each part was done to ensure comparability with both the social and computer conditions, characterized by a consistent "pacer" (same experimenter, same computer). Fourth, both melodies were simple with a clear development and a predictable ending point to ensure appropriate action from the participant when required.

4.4.2.3 Experimental design

We manipulated 3 factors: 2 levels of music (present, absent), 2 levels of social interaction (present, absent) and 2 levels of sentence context variability (same, different).

Music context and social interaction were both evaluated as between-subject factors. Every participant was semi-randomly assigned to one of four conditions: music and social interaction (M+, S+; N = 20, 10F, mean age 24.40 years, SD 2.04), non-music and social interaction (M-, S+; N = 20, 10F, mean age 24.30 years, SD 2.23), music and non-social interaction (M+, S-; N = 20, 10F, mean age 24.85 years, SD 3.12), and lastly non-music and non-social interaction (M-, S-; N = 20, 10F, mean age 25.90 years, SD 2.83). There was no age difference between the groups [all ps > .089]. The four groups were additionally balanced in terms of their musical background, defined in terms of years of musical practice prior to the participation in the study (mean number of years of instrument playing = 4.99, SD 6.37; mean number of years of singing and/or dancing = 1.94, SD 4.26; all ps > .210).

Half of the objects (N = 10) occurred repetitively within the same sentence context (sSC –same sentence context). For example, the image representing "the cow" was always the correct ending for the same sentence context "the wolf bites". The other half of the objects (N = 10) was presented at each repetition within a different sentence context (dSC – different sentence context). For example, the image representing "the egg" could follow in sentence contexts such as "the woman cuts", "the boy eats", etc. The alternation between sSC and dSC checkerboards was randomized. Although each sentence was repeated 9 times, the actual number of exposures to each pseudo-word was dependent on the number of correct responses given by each participant, as a pseudo-word was presented only in case of the correct object identification.

4.4.2.4 Task and Experimental Procedure

The experiment consisted of three parts: Practice trials, learning phase, testing phase. Stimuli were presented using a desktop computer running Presentation 16.0 (Neurobehavioral Systems, Albany, USA). Two standard wheel mice (Premium Optical Wheel Mouse, Logitech, Morges, Switzerland) were connected to the same Windows computer and used as response devices.

Musical stimuli were presented via a stereo speaker system (LS21 2.1, Logitech, Morges, Switzerland). The task specifics are described below.

4.4.2.4.1 Practice trials and learning phase

Participants were first presented with detailed written instructions and performed a block of 10 practice trials to familiarize themselves with the task requirements. In all conditions, the task of the participant was to find the correct object for a given sentence context amongst the images on the checkerboards. Each trial began with the presentation of a fixation cross (500 ms), followed by a checkerboard containing 9 images. In each checkerboard, a red frame appeared around the image representing the subject of the sentence context, followed by a second red frame around the image representing the verb of the sentence context. When both elements were marked with a red frame, the participants could give their answer by selecting an object fitting the sentence context from the remaining 7 images on the checkerboard.

For participants assigned to the **social condition**, the subject and verb of the target sentence were selected on the checkerboard by the experimenter. Participants assigned to the **non-social condition** (both M+ and M-) had the sentence context selected by the computer program.

In the **M- condition**, the red frame around the "subject" appeared with a variable stimulus onset asynchrony (range: 2.630 – 4.650 sec); the red frame around the "verb" followed after 600 ms. In the **M+ condition**, a melody started playing when the checkerboard appeared; the "subject" red frame was highlighted at the end of the melody. The duration of the melodies was comparable to the stimulus onset asynchrony of the M- condition (range: 2.637 – 4.642 sec). In order to make all the conditions maximally comparable, the duration of the jitters and melodies were based on the experimenter's times in a previous study using the same paradigm (Verga & Kotz, in prep.). The experimenter was the same in both experiments.

There was no time limit for participants to answer. In all conditions, if a correct answer was given, the selected image was substituted by a pseudo-word (black capital letters over white background, Arial, 40 pt) providing the “Italian name” of the object. The pseudo-words remained on the screen for 1000 ms. If an incorrect response was given, no “Italian name” was displayed and the following trial began immediately.

After the training, participants performed the learning phase. The procedure of the learning phase was identical to the training phase. 180 trials (20 objects x 9 repetitions) were presented in total during the experiment.

4.4.2.4.2 Testing Phase

At the end of the learning phase, a behavioral testing phase took place to evaluate whether pseudo-words presented during the learning phase had been mapped to the corresponding objects. In this task, participants were presented with novel sentence contexts (that is, combinations of pictures representing a subject and a verb that had not been seen together before), followed by three of the pseudo-words (“Italian words”) participants had learned during the learning phase. Participants were asked to select the “Italian word” that matched a given sentence context. All trials contained one correct and two incorrect options.

4.4.2.5 Data analysis

Statistical analyses of behavioral data were performed using MATLAB R2013a (The Mathworks Inc., Natick, USA) and IBM SPSS Statistics 18 (IBM Corporation, New York, USA).

Behavioral data were first corrected for outliers. Trials with response times exceeding the mean ± 2 SDs were excluded from further analysis (mean rejected trials across participants = 4.32 %).

For the learning phase, response times were calculated as the time delay between the appearance of the “verb” image and the participant’s answer. Accuracy scores (proportion of correct responses in total), response times for correct responses and their SDs were calculated for each repetition of the object, for each participant. To evaluate the degree of temporal coordination of the participant during the learning phase, we used the following measures: First, SDs of response times were employed as an index of the stability of participants’ performance. We additionally used the coefficient of variation (CV) as an index of variability independent of response speed, to allow for a direct comparison between the different conditions. Further, we calculated the lag-0 and lag-1 cross correlation (cc) coefficients between the inter-trial-intervals produced by the participants and those produced by the experimenter (S+ conditions) or computer (S- conditions). More specifically, the cross correlation at lag-0 indicated how much the behavior of the participant in one trial was temporally related to the behavior of their partner (the experimenter/computer) in the same trial. Cross-correlations at lag-1 indicated whether the behavior of the experimenter/computer was related to the participant’s behavior in the following trial. To account for the difference in the variability of trial presentation in the different conditions, we conducted separate ANCOVAs on the variables of interest using the SDs of the experimenter’s/computer’s response times as covariates during the learning phase. We did not use this covariate in the cross-correlation analyses as SDs account for the variability in the computer/experimenter RTs series, in which the correlation coefficients are calculated.

For the testing phase, response times were calculated as the time delay between the appearance of the three alternative pseudo-words and the participant’s response. Accuracy scores were defined as the proportion of correct responses out of the total number of responses. We used the number of exposures during the learning phase as a covariate. This number took into account the mean number of times pictures were repeated during the learning phase, ranging from a minimum of 0 (no correct responses) to a maximum of 9 times (no errors).

When the assumption of sphericity was not met, a Greenhouse-Geisser correction was applied to the degrees of freedom. Two-tailed t-tests and simple effect analyses were employed to compare individual experimental conditions and to resolve interactions. We used an alpha level of .05 to ascertain significance for all statistical tests, and applied a Bonferroni correction in post-hoc tests to control for multiple comparisons.

4.4.3. Results

4.4.3.1 Learning Phase

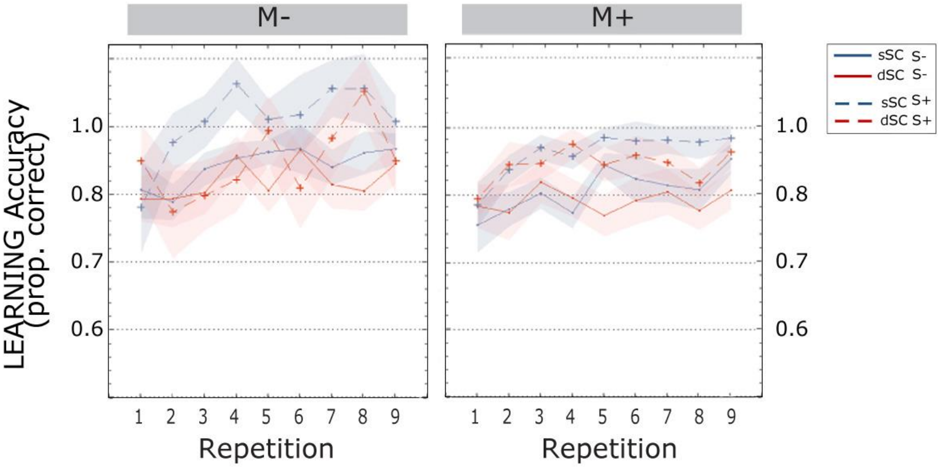
Participants responded with an average accuracy of 93% correct. A 2x2x3 repeated measures ANCOVA was conducted on accuracy scores with the between factors music context (M+ vs. M-) and social context (S+ vs. S-), the within factors sentence context (dSC vs sSC) and repetition (11 repetitions), and SDs of presentation times (experimenter, computer) as covariates to account for differences in variability across conditions.

Participants' **accuracy** increased during the learning phase [linear trend, F(4.569, 333.552) = 5.798, p = .000, η_p^2 = .074]. Words encoded in same sentence contexts (sSC, M = .954, SEM = .008) elicited higher accuracy than words encoded in different sentence contexts (dSC, M = .925, SEM = .009) [F(1,73) = 14.782, p = .000, η_p^2 = .168]. There were no other significant effects or interactions (all ps > .074; **Figure 18**).

Response times decreased over the course of the learning phase [linear trend, F(3.046, 219.321) = 34.332, p = .000, η_p^2 = .323]. Words encoded in

Figure 18 – Learning phase accuracy scores.

Accuracy scores are plotted as a function of item repetitions and controlled for time variability in sentence context presentation. The area subtended by the shadows represents standard errors of the mean. Abbreviations: M- = music context absent; M+ = music context present; S- = non social interaction; S+ = social interaction; dSC = different sentence context; sSC = same sentence context.

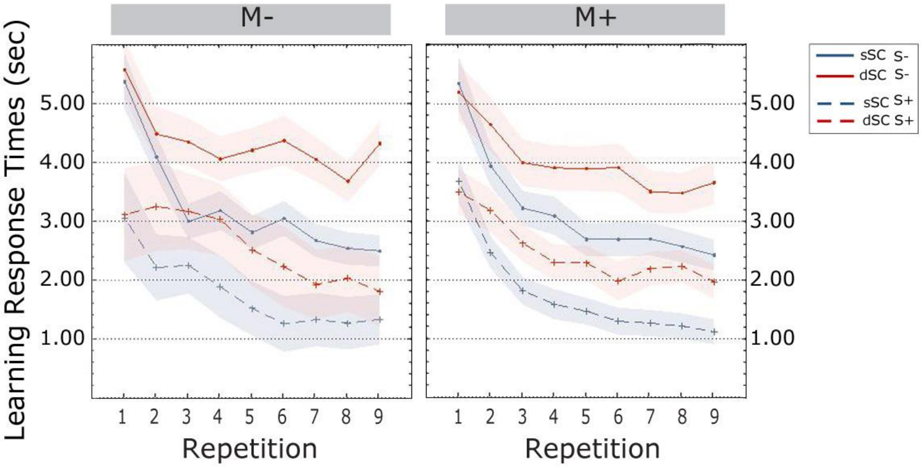


different sentence contexts elicited slower response times (dSC, M = 3.339, SEM = .139) compared to words encoded in same sentence contexts (sSC, M = 2.487, SEM = .107) [F(1,72) = 73.839, p = .000, $\eta_p^2 = .506$]. The interaction between repetitions and sentence context was significant: Bonferroni corrected post-hoc tests revealed no difference between sSC and dSC words at the first repetition (p = .863); however, response times for the two conditions started to differ already with the second repetition, with sSC being significantly faster than dSC during the entire learning phase (all ps < .001).

Participants trained socially (S+, M = 2.325, SEM = .174) were significantly faster than participants trained non-socially (S-, M = 3.485, SEM = .174) [F(1,72) = 11.471, p = .001, $\eta_p^2 = .137$]. There were no other effects or significant interactions (all ps > .103**; Figure 19**).

Figure 19 – Learning phase response times.

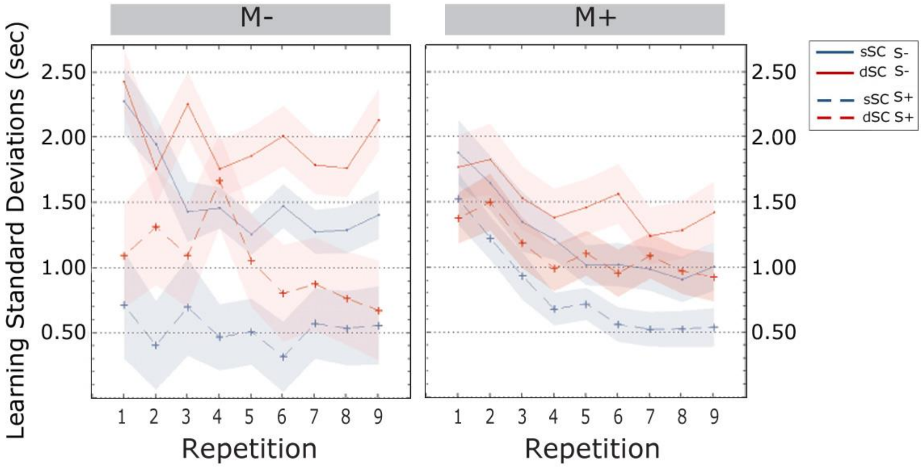
RTs are plotted as a function of item repetitions and controlled for time variability in sentence context presentation. The area subtended by the shadows represents standard errors of the mean. Abbreviations: M- = music context absent; M+ = music context present; S- = non-social interaction; S+ = social interaction; dSC = different sentence context; sSC = same sentence context.



Standard deviations of the response times decreased over the course of the learning phase [linear trend, F(5.490, 395.256) = 3.625, p = .002, $\eta_p^2 = .048$]. Bonferroni corrected post-hoc tests revealed that variability was significantly different between the first and the second item repetition (p = .000), between the second and the third (p = .019) and between the fourth and the fifth repetition (p = .020). There was no difference between the other transitions from one repetition to the next (all ps > .796). Further, standard deviations for the responses to sSC words (M = 1.019, SEM = .051) were smaller than those to dSC word (M = 1.402, SEM = .071) [F(1,72) = 35.722, p = .000, $\eta_p^2 = .332$]. Additionally, participants trained in a social interactive context (S+, M = .869, SEM = .123) were less variable than participants trained non socially (S-, M = 1.552, SEM = .130) [F(1,72) = 9.347, p = .000, $\eta_p^2 = .115$]. There were no further effects and no interactions (all ps > .113; **Figure 20**).

Figure 20 – Learning phase standard deviations of response times.

SDs are plotted as a function of item repetitions and controlled for time variability in sentence context presentation. The area subtended by the shadows represents standard errors of the mean. Abbreviations: M- = music context absent; M+ = music context present; S- = non-social interaction; S+ = social interaction; dSC = different sentence context; sSC = same sentence context.

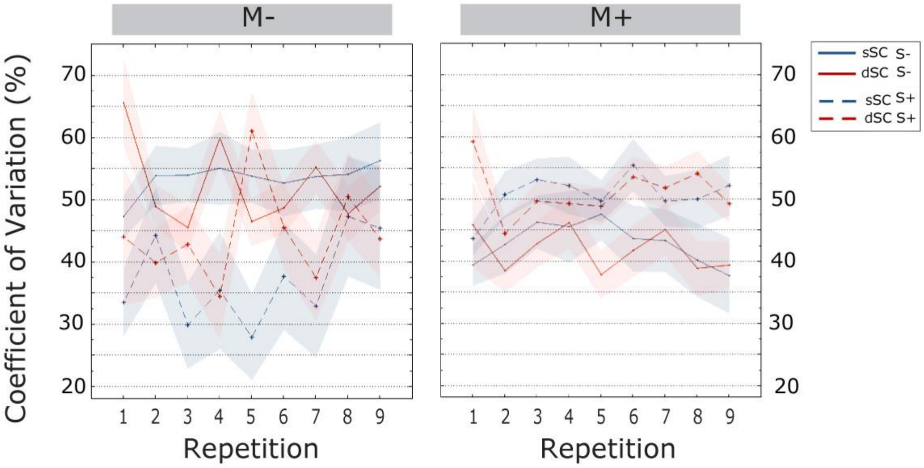


The **Coefficient of Variation** (CV) increased over the course of item repetitions [linear trend, F(6.355, 457.583) = 2.813, η_p^2 = .038]. Bonferroni corrected post-hoc tests revealed that the coefficient of variation was significantly lower in the first item repetition as compared to all subsequent repetitions (all ps < .033); further, in all repetitions except the third and seventh, the CV was lower than the last one (all ps < .038).

Additionally, we observed an interaction between music context and social interaction [F(1,72) = 12.173, p = .000, η_p^2 = .145]. Therefore, a simple effect analysis was carried out. This analysis revealed that participants trained non-socially had significantly more stable performances when doing the task with music (M+, M = .373, SEM = .026) than without (M-, M = .478, SEM = .026) [F(1,72) = 13.681, p = .000, η_p^2 = .160]. In socially-trained participants, we

Figure 21 – Learning phase coefficient of variation.

CV is plotted as a function of item repetitions and controlled for time variability in sentence context presentation. The area subtended by the shadows represents standard errors of the mean. Abbreviations: M- = music context absent; M+ = music context present; S- = non-social interaction; S+ = social interaction; dSC = different sentence context; sSC = same sentence context.



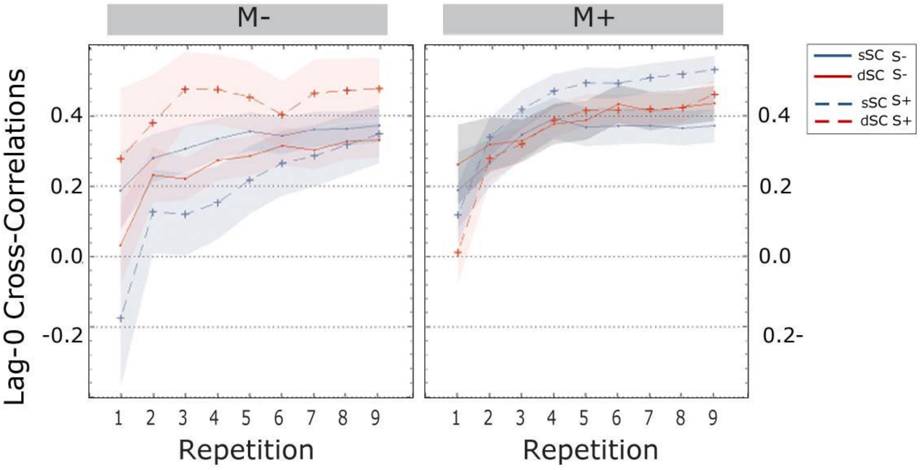
observed the opposite effect, though this was only marginally significant: Participants performing the music task had significantly higher values of CV (M = .459, SEM = .020) as compared to participants doing the task without musc (M = .357, SEM = .042) [F(1,72) = 3.825, p = .054, $\eta_p^2 = .050$]. There were no other significant effects or interactions (all ps > .099; **Figure 21**).

The **cross-correlations at lag-0** revealed a main effect of repetition [linear trend, F(1.897, 142.252) = 70.639, p = .000, $\eta_p^2 = .485$]; more specifically, Bonferroni corrected post-hoc tests revealed a significant increase from one repetition to the next (all ps <.001) except for repetitions 4, 5 and 6 (all ps > .083).

The difference between social groups was significant [F(1,75) = 8.044, p = .006, $\eta_p^2 = .097$]; indeed, participants trained socially had significantly higher

Figure 22 – Learning phase lag-0 cross-correlations coefficients.

Lag-0 cc are plotted as a function of item repetitions. The area subtended by the shadows represents standard errors of the mean. Abbreviations: M- = music context absent; M+ = music context present; S- = non-social interaction; S+ = social interaction; dSC = different sentence context; sSC = same sentence context.



lag-0 cc values (S+, M = .387, SD = .025) compared to participants trained non-socially (S-, M = .286, SEM = .025).

Further, the three-way interaction between sentence context, social interaction and music context reached significance [F(1,75) = 11.435, p = .001, $\eta_p^2 = .132$]. A follow-up simple effects analysis revealed that when participants were trained in a musical context, there were no differences if they were trained with a partner or without [F(1,75) = 1.260, p = .265, $\eta_p^2 = .017$], nor were there differences for sSC compared to dSC words [F(1,75) = .017, p = .897, $\eta_p^2 = .000$]. However, when learning without music, participants trained socially displayed significantly higher lag-0 correlations for dSC words compared to sSC words [dSC, M = .471, SEM = .044; sSC, M = .324, SEM = .043; F(1,75) = 9.323, p = .003, $\eta_p^2 = 111$]. There was no difference between sSC and dSC words for participants trained non-socially without music [$F(1,75) = .291$, $p = .591$, $\eta_p^2 = .004$].

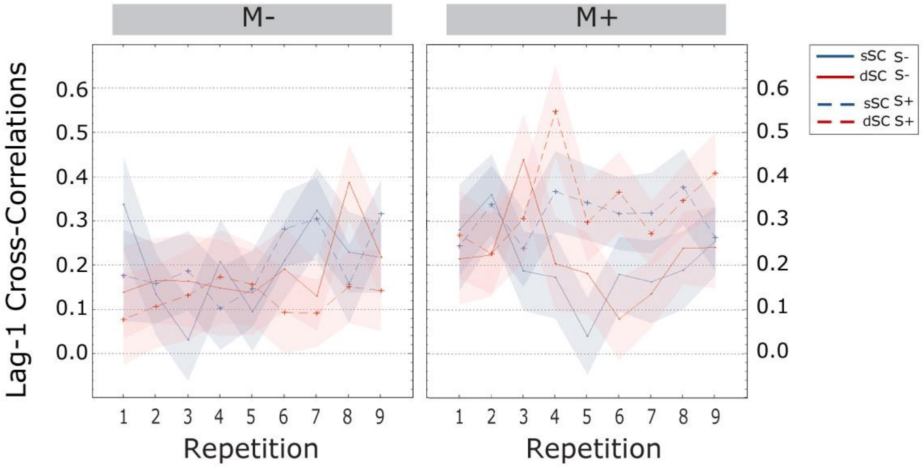
The three-way interaction between repetition, social interaction and music context was also significant [$F(1.897,142.252) = 4.120$, $p = .020$, $\eta_p^2 = .052$], therefore a simple effects analysis was carried out. This analysis revealed that when learning without music, participants in the S+ group had from the very beginning higher lag-0 cc (M = .245, SEM = .062) than participants trained non-socially (S-, M = .031, SEM = .061) [$F(1,75) = 6.035$, $p = .016$, $\eta_p^2 = .074$]. There was no difference when participants were trained with music at the first repetition [$F(1,75) = 1.698$, $p = .196$, $\eta_p^2 = .022$]. There was no difference between the two groups (S+ and S-) in either music condition (M+, M-) in repetitions 2, 3 and 4. Starting from the fifth repetition, participants learning without music became significantly more coordinated when trained with a social partner compared to a computer. This effect was then continuous until the end of the experiment (for all repetitions $p < .025$). The same significant difference was found in the musically trained group, but only starting from the second to last repetition (for repetitions 8 and 9 ps < .044). There were no other significant effects or interactions between factors (all ps >.120; **Figure 22**).

The **cross-correlations at lag-1** were significantly higher for participants trained with music (M+, M = .167, SEM = -017) than without (M-, M = .078, SEM = .017) [$F(1,72) = 13.572$, $p = .000$, $\eta_p^2 = .159$]. Further, the interaction between social interaction and music context was significant [$F(1,72) = 8.676$, $p = .004$, $\eta_p^2 = .108$], therefore a simple effects analysis was carried out. This analysis revealed no difference between participants trained socially or non-socially when learning without music [$F(1,72) = .671$, $p = .415$, $\eta_p^2 = .009$]. However, participants trained with music had significantly higher lag-1 correlations when playing with a partner (M = .224, SEM = .024) compared to a computer (M= .110, SEM = .024) [$F(1,72) = 11.672$, $p = .000$, $\eta_p^2 = .137$; **Figure 23**].

To summarize, learning effects emerged during the task with a progressive increase in accuracy and temporal coordination (lag-0 cc) and a decrease in response times. Overall, words encoded in a consistent sentence context were recognized faster and more accurately than words encoded in a different context. Participants trained socially were significantly faster, less variable (SDs) and more temporally coordinated (lag-0 cc) than participants trained non-socially. In the no-music condition, lag-0 cross correlations were significantly higher for social participants exposed to dSC words. However, in the music condition no differences were observed. Variability independent of speed (CV) was lower for participants who trained non-socially with music than without; participants playing with an experimenter were instead more stable without music. Lag-1 cross-correlations were higher for participants trained with music, especially when playing the game with a partner.

Figure 23 – Learning phase lag-1 cross-correlations coefficients.

Lag-1 ccs are plotted as a function of item repetitions. The area subtended by the shadows represents standard errors of the mean. Abbreviations: M- = music context absent; M+ = music context present; S- = non-social interaction; S+ = social interaction; dSC = different sentence context; sSC = same sentence context.



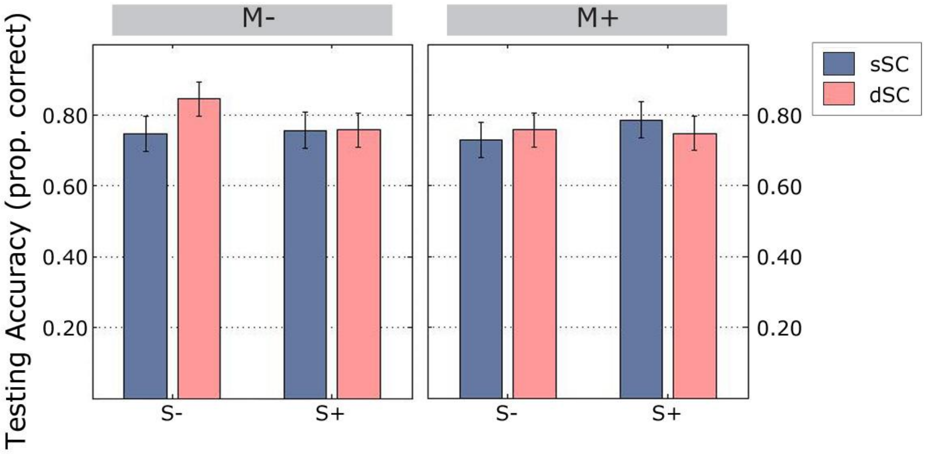
4.4.3.2 Testing phase

Separate 2x2x2 ANCOVAs were conducted on accuracy scores and response times to evaluate the impact of the experimental manipulations (music context, M+ vs. M-; social context, S+ vs. S; sentence context, sSC vs. dSC) while accounting for the number of exposures to the pseudo-word during the learning phase.

Overall, participants performed at an accuracy level of 77%. We observed a significant interaction between sentence context and social interaction [F(1,75) = 4.605, p = .035, η_p^2 = .058], therefore a simple effects analysis was carried out. This showed that there was no difference between sSC and dSC words in the group of participants trained socially [F(1,75) = .465, p = .497, η_p^2 =.006].

Figure 24 – Testing phase accuracy scores.

Accuracy scores controlled for mean number of repetitions during learning. Vertical lines represent standard errors of the mean. Abbreviations: M- = music context absent; M+ = music context present; S- = non-social interaction; S+ = social interaction; dSC = different sentence context; sSC = same sentence context.



However, participants trained in the S- condition correctly identified more dSC (M = .801, SD = .209) than sSC (M = .739, SD = .233) words, [F(1,75) = 5.536, p = .021, $\eta_p^2 = .069$]. There were no other significant interactions (all ps > .151) and no significant main effects (all ps > .204).

Response times during the testing were not significantly different between conditions when controlling for mean repetitions during the learning phase (all ps > .193).

In summary, during the testing phase, participants trained non-socially remembered more words originally encoded in different sentence contexts.

4.4.4. Discussion

The aim of the current study was to investigate the impact of music and social interaction on adult word learning. Both types of context have been hypothesized to enhance attention towards relevant information in the learning environment (that is, the referent for a new word), by exerting a unidirectional (music) or bidirectional (social interaction) temporal influence on the learner. To address whether this difference impacts the way new words are learned, we implemented a game set-up, in which participants learned new words embedded in sentence contexts with different degrees of variability. Our results show that participants were significantly faster, less variable and more temporally coordinated when learning with a partner, than when participants were trained non-socially. When learning without music, participants trained socially displayed better coordination during variable (“difficult”) context trials compared to consistent (“easy”) context trials. However, coordination with music, especially when playing with a partner, tended to “shift” from one trial to the next. Variability, when accounting for differences in response times, was lower for participants learning non-socially in the music condition. Finally, in the testing phase, participants trained non-socially remembered more words originally presented in different sentence contexts, although words repeated in a consistent context represented an easier condition (confirmed by the faster reaction times and higher accuracy in this condition during learning). While these results are in line with previous evidence of spontaneous temporal coordination during social interaction, they also provide a significant advance for research on communication and word learning in adults; indeed, they suggest that not only are adult learners influenced by the presence of a social partner, but also that this influence is different from the one exerted by other sources, such as music.

The results presented here support previous literature showing that temporal coordination spontaneously emerges during social exchanges (Demos et al., 2012; Richardson et al., 2007; Yun et al., 2012). Indeed, participants performing the task with a social partner were faster, less variable, and more temporally coordinated with the experimenter than participants performing the task with a computer. Temporal coordination with music had a weaker effect as compared to social interaction, as participants coordinated their behavior with the music stimuli immediately preceding the one they were listening to. These results can be interpreted within the framework of coupled oscillators (Demos et al., 2012; Dumas et al., 2010; M. Wilson & Wilson, 2005). In brief, this hypothesis proposes that since human movements tend to be rhythmic, two people performing a joint task are not dissimilar from other systems displaying periodic variations in time. As a consequence, interacting human dyads respond to the same dynamics as other oscillators; that is, they reciprocally influence each other in order to reach an equilibrium (Richardson et al., 2007; see also Kelso, 1997). Music, on the other hand, represents a unidirectional influence. In the present study, participants coordinated with the temporal regularities of the music, but the lack of reciprocal adaptation reduced the extent of the coordination. While this result seems in contrast with evidence that has consistently shown a strong effect of music on temporal coordination, it must be noted that, in most previous studies investigating sensori-motor synchronization, participants have been explicitly required to coordinate with continuously playing sequences (Pecenka, Engel, & Keller, 2013; Repp & Su, 2013). Instead, we wanted to exploit *spontaneous* coordination with a *temporally defined* musical excerpt (that is, the musical sequence was finite for each trial and participants were required to take action at the end of the sequence, not during it), in order to maximize the music's potential to drive the learner's attention to a specific point in time. Results from the condition in which both music and a partner were present at the same time further corroborate this interpretation: Music and social interaction may be responsible for different forms of coordination, due, in turn, to different underlying mechanisms. Indeed, participants learning socially are significantly more variable in their responses when learning with music, while the opposite is true for participants learning alone (we observed less variable performances with music). This increased behavioral uncertainty likely depends on the different influences stemming from the two sources. While without music there is only one source of information (the experimenter), music introduces a second set of coordinative cues; since the two sources exert different influences (unidirectional versus bidirectional), there may be uncertainty as to what one should coordinate to. In turn, this uncertainty is behaviorally reflected in an increased response variability. However, this uncertainty is likely transient; increased coordination with the experimenter (compared to the computer) when music was present, emerged only towards the end of the learning phase, much later than without music. Furthermore, this coordination with music was maximal between responses in one trial and the music excerpt of the preceding trial, but not with the music in the trial participants were responding to; in other words, participants' tended to have response patterns which reflected the duration of the previous musical stimulus, but not the one they were answering to. Another explanation for these results may be that in the current task, music was employed concurrently with another high-level cognitive task (identifying a sentence on the checkerboard). Despite the relative simplicity of the musical stimuli that we employed, the combination of music and task demands may have been too challenging for music to actually facilitate the performance (Kang & Williamson, 2013). However, the lack of difference in response accuracy or reaction times in the music and non-music conditions tends to rule out this possibility.

So far, the results of the learning phase suggest that temporal coordination to music and a social partner have different characteristics, possibly reflecting different underlying mechanisms. But what are the implications for word learning? Both music and social interaction have been claimed to facilitate word learning and memory (De Groot, 2006; Ferreri et al., 2014, 2013; Jeong et al., 2010; Ludke et al., 2014; Rainey & Larsen, 2002; Verga & Kotz, 2013); several accounts explain this effect as the result of the easiness – for these stimuli – to allow predictions on the upcoming events and allocate one’s attention accordingly (Gogate et al., 2000; Lagarde & Kelso, 2006; Rader & Zukow-Goldring, 2012; Rolf et al., 2009; Schmidt-Kassow et al., 2013). The data presented here, however, suggest that the behavioral adjustments participants make may be based on different kinds of predictions. In the case of music, predictions are based on the temporal structure of the stimulus (unidirectional influence), while in the case of a partner they rely on the ability to infer the other person’s intention (bidirectional influence; Frith & Frith, 2006, 2012). This allows the creation of a “psychological common ground”, in which the transmission of information is facilitated (Csibra & Gergely, 2009; Tomasello, 2000). In this shared psychological space, the increased temporal coordination observed in this study may reflect a strategy that a knowledgeable partner uses to direct the learner's attention towards the correct referent for a new verbal label (Pereira et al., 2008). Thus, the attention of the learner is focused on the target referent, consequently facilitating the mapping of a new word onto its meaning. This account predicts that temporal coordination with a knowledgeable partner should be better when the learner does not know *a priori* where the target referent may occur. In this situation, the adult learner is similar to a child learning its first words and faced with a constantly changing environment, in which multiple referents are present. Our results show that, indeed, temporal coordination with the experimenter was higher in this contextual condition, replicating our previous findings (Verga & Kotz, in prep.). However, no differences were found between music and non-music conditions in relation to the variability of the context that words were embedded in. An interpretation of this result is that a shared psychological space – behaviorally reflected in the temporal coordination with a partner – is used by adult learners to identify a referent for a new word, when it cannot be extracted by the context of the word presentation alone. That is, participants "disengage" from social interaction if they can identify a referent by themselves. Instead, the presence of music overrules contextual diversity, as participants maintain the same pattern of coordination independently from the characteristics of a word presentation. This result is somehow in-between the two opposing accounts of the adult learner, one suggesting that adults are entirely self-sufficient learners (Pickering & Garrod, 2004; Stephens et al., 2010) and the other suggesting a critical role for others in shaping cognitive activity (Ciaramidaro et al., 2014; Schilbach, 2014; Schilbach et al., 2013; Sebastiani et al., 2014); indeed, these results suggest that the presence of another person is used *when needed*. While our results indeed confirm that music and social interaction may drive attention in different ways, the question remains open as to which strategy may be more relevant to successfully learning new words. An important implication of these results concerns situations in which music and social interaction are present at the same time, especially for tasks requiring coordination to either one of the two stimuli. Music therapy represents an important example of this situation. In addition to its positive effect on mood and arousal (Sarkamo et al., 2008), music is often employed to provide the patient with a temporal structure to facilitate her/his performance (Stahl, Henseler, Turner, Geyer, & Kotz, 2013; Stahl, Kotz, Henseler, Turner, & Geyer, 2011), while at the same time a therapist needs to be present with the patient (Norton, Zipse, Marchina, & Schlaug, 2009). The competition observed in this study between music and a social partner as coordinative tools suggest that their respective roles should be further investigated in these types of settings.

Quite surprisingly, during the testing phase, participants that were trained non-socially correctly identified more words when they had originally been presented in variable sentence contexts (as opposed to consistent sentence contexts), while no differences were observed either in the social group or in the music groups. In general, an advantage of words repeated at each occurrence in a different context is to be expected, as every time the same word is encountered in a different context, different contextual cues accumulate and enrich the representation of the target referent and its association with the new word (Adelman et al., 2006; Lohnas, Polyn, & Kahana, 2011; Verkoeijen, Rikers, & Schmidt, 2004). Nevertheless, according to the hypothesis that a social partner and music may help the learner in directing attention toward the target (although through different mechanisms), an advantage of music and social interaction over simple computer learning should be expected. We provide two possible explanations for these results: First, while learning new words from a computer interface and testing participants with a computer interface is consistent, participants who learned with social interaction and/or with music may have been disadvantaged as they experienced a contextual inconsistency between the learning and the testing phase. Indeed, consistency between learning and testing environments has been suggested to facilitate recall (Godden & Baddeley, 1975; Polyn, Norman, & Kahana, 2009). This hypothesis, known as the "transfer appropriate processing" theory, states that the strength of a memory trace (that is, the ease of its retrieval) depends on the type of encoding compared to the type of retrieval (Stein, 1978; Tulving, 1979); if the form of encoding is congruent with the type of testing, retrieval is facilitated. In this study, the social and the music group faced an incongruity between the learning phase and the retrieval phase, which was always conducted by participants alone and without music. Instead, the non-social groups were exposed to the same type of encoding and testing (both alone and without music). An explanation based on incongruence between the type of encoding and the type of testing has been suggested in other learning studies; for example, Peterson and Thaut (Peterson & Thaut, 2007) found no behavioral advantage for sung compared to spoken word lists in an explicit verbal learning task, in which words were sung during learning and spoken during the recall phase. However, a behavioral advantage for sung stimuli emerged when participants were instructed to sing back during the recall phase (Thaut, Peterson, Sena, & McIntosh, 2008; Thaut, Peterson, McIntosh, & Hoemberg, 2014; for a review see Ferreri & Verga, in prep.). Further investigation is required to clarify this aspect, by testing participants in the same condition they were trained in. Results in this direction would have important implications in terms of the extent to which acquired knowledge may be generalized to different contexts. If the context of word acquisition needs to be the same at retrieval, this would have little facilitation in some conditions, for example, if music needs to be present every time the new word is used, it would not be particularly helpful. The case of social interaction somehow represents an exception, as words are often (although not always) learned with someone (for example, in first language learning) and used to communicate with someone. Hence, in this condition, results favoring the transfer appropriate theory would not be so problematic.

Second, our testing phase took place immediately after the learning phase, and therefore we did not consider consolidation effects that have been deemed important for word learning in both children and adults (for example Henderson, Weighall, Brown, & Gaskell, 2013). Social context has been proven to significantly bias the formation of new memories. For example, in a study by

Straube and colleagues (B. Straube, Green, Chatterjee, & Kircher, 2010), participants watched video clips of an actor speaking to them directly or to a third person. Source memory (the memory of the context a sentence was heard in) was significantly biased by social interaction, as participants tended to report that the actor was talking to them even if he was not. In our experiment, the testing phase took place immediately after encoding and it did not provide information concerning possible long-term mnemonic effects, which critically depend upon consolidation processes (Walker & Stickgold, 2004). As the efficacy of consolidation depends on several factors, among which sleep seems to play a particularly pivotal role (Atherton, Nobre, Zeman, & Butler, 2014; Diekelmann & Born, 2007; Lewis, 2014; Siegel, 2001; Stickgold, 2005), a possible way to test long-term effects of social interaction may be by testing retrieval at delayed time points after the learning phase has taken place.

To conclude, the current study aimed at investigating the respective roles of music and social interaction as possible facilitators of word learning in healthy adult speakers. We found that social interaction, more than music, improves temporal coordination in a verbal learning task. Further, music and social interaction provide different types of influence (unidirectional versus bidirectional) that do not combine together easily, as the presence of social interaction and music at the same time hinders coordination. Crucially, the quality of coordination with the human partner (but not with music) is intertwined with the attentional demands of the task at hand; coordination is higher when it is difficult to find a new word's referent. Taken together, these results support the notion that music elicits a different form of temporal coordination from the one observed in interacting dyads, whose behavior is compatible with coupled oscillators. This result has important implications for situations in which music and social interaction are present at the same time, such as many forms of music therapy. Although different, these forms of coordination equally impact word learning, as seen in the testing phase immediately following the task. This result calls for further study to elucidate the extent to which the context of learning influences performance during retrieval and how the latter may be influenced by consolidation processes.

4.5 Neural correlates of social word learning[19]

Abstract

In previous studies we have shown that social interaction modulates adult word learning in concert with the characteristics of the sentence context a word is embedded in. When the context a word is presented in changes at each occurrence, the learner does not know a priori what the target referent is. Similarly to children, adults in this situation may benefit from the presence of a social partner directing the learner's attention towards the correct word referent in the visual environment. If this is the case, at the neural level we would expect the activity of the circuit underlying visuo-spatial attention (involving the middle frontal gyrus, angular gyrus and visual cortices) to be enhanced during interactive learning, leading to increased activation in areas involved in the task at hand (that is, word learning – recruiting the middle and inferior temporal gyrus, inferior frontal gyrus, temporo-parietal junction, as well as subcortical structures). In the current study, we investigated this hypothesis with a comprehensive set of analyses, involving a mass univariate GLM analysis and an Independent Component Analysis (ICA). Further, we specifically focused on the connectivity pattern in one of the relevant networks, namely the fronto-parietal attentional reorienting network, by performing Dynamic Causal Modelling (DCM). Our results suggest that social interaction influences the processes underlying word learning by modulating activity of task-related areas. More specifically, we propose that the presence of a social partner modulates the activity of the network involved in visuo-spatial attention. Further, we suggest that this modulation is dependent upon task specifics: If the task is easy enough for the learner (in other words, if the sentence context conveys enough information), social interaction does not provide an advantage.

---

[19] This study is based on the paper: "Neural correlates of social word learning", by Verga & Kotz (in preparation).

These results provide evidence that social interaction may influence visuo-spatial attention, facilitating word learning in a second language.

4.5.1 Introduction

Language learning is typically a social activity, often characterized by an exchange of information between a learner and a knowledgeable partner. Nevertheless, our current knowledge of this process is limited to what happens in the learner's brain during individual learning. In fact, only recently has the impact of social interaction on cognition become of interest to neuroimaging researchers. A first attempt to study social word learning in adults has been made by Jeong and colleagues (2010), who investigated word learning in adult Japanese speakers learning Korean words. Participants were presented with movie clips depicting either text-based learning contexts (new words were spoken by a person holding their written translation) or situation-based contexts (new words were exchanged during real life interactions between two actors). A post-learning functional magnetic resonance imaging (fMRI) test revealed that the right supramarginal gyrus (rSMG) was involved in retrieval of L2 words encoded in a social setting (Jeong et al., 2010). This result is in line with several studies on social cognition that have consistently reported activations for social stimuli in areas inside and around the right temporo-parietal junction (TPJ), a brain region at the boundary of the temporal and parietal lobe roughly corresponding to BA39 (Carter & Huettel, 2013; Jeong et al., 2010). While activations in the parietal lobe are extremely consistent across various types of social cognition studies, the question that remains is: What role does the parietal cortex play during social interactions?

This question stems from evidence that the right parietal cortex was recruited in experiments investigating low-level cognitive processes (Carter & Huettel, 2013; Elman, Rosner, Cohn-Sheehy, Cerreta, & Shimamura, 2013) as well as social interaction; in particular, there is strong evidence for a high degree of overlap in and around the right TPJ during social cognition and visuo-spatial attention tasks (Decety & Lamm, 2007). In the latter case, the role played by the right parietal cortex in visuo-spatial attention is to work as a junction between superior frontal regions, such as the middle frontal gyrus and the frontal eye fields (BA8; Corbetta, Patel, & Shulman, 2008; Corbetta & Shulman, 2002) and the visual cortex (Constantinidis, Bucci, & Rugg, 2013; Macaluso, Frith, & Driver, 2000; Saalmann, Pigarev, & Vidyasagar, 2007; Verghese, Kolbe, Anderson, Egan, & Vidyasagar, 2014), with the possible function of re-directing attention towards new targets (Corbetta, Kincade, Ollinger, McAvoy, & Shulman, 2000). More specifically, superior frontal areas such as middle frontal gyrus (Thiel, Zilles, & Fink, 2004) facilitate the activation of the task-relevant visual cortices via a top-down connection from the angular gyrus and superior visual cortices to the striate cortex (Chambers, Payne, Stokes, & Mattingley, 2004; Horwitz, Rumsey, & Donohue, 1998), ultimately facilitating the processing of stimuli at attended locations (Bressler, Tang, Sylvester, Shulman, & Corbetta, 2008; Thiel et al., 2004).

Is it possible that this function also applies to social language learning? And why should this be the case? Studies on first language learning suggest that, indeed, sharing visual attention with a caregiver is a sine qua non condition for successful verbal learning (Kuhl, Tsao, & Liu, 2003; Waxman & Gelman, 2009), as a knowledgeable partner may direct the learner's attention towards the correct referent from among many possible targets (Dominey & Dodane, 2004; Tomasello, 2000; Verga & Kotz, 2013); thus, without explicit instructions, children can easily learn new words from context. While second language (L2) learning certainly has some points of divergence from first language (L1) learning[20], even adults can acquire new words incidentally from context (Laufer

---

[20] Second language (L2) is usually learnt via explicit formal training as compared to a first language (L1), which is usually acquired effortlessly without explicit instructions (see for example Abutalebi et al., 2008). However, the case of word learning represents a bridge between L1 and L2 learning, as even new words in L2 can be acquired without explicit training.

& Hulstijn, 2001; Nagy et al., 1987; Rodríguez-Fornells et al., 2009; Swanborn & De Glopper, 1999); however, in this situation, multiple possible referents for each meaning are available. To solve this problem, adult learners capitalize on cues provided by the context the word is presented in; this sentence context helps the learner to constrain a set of possible meanings (Mestres-Missé, Càmara, Rodriguez-Fornells, Rotte, & Münte, 2008). However, the impact of other variables – such as the presence of another person – has not been clarified as of yet. Does a knowledgeable partner facilitate adult learners as she/he does with children, by directing their attention towards a particular referent? And if so, does it always happen, or does this depend on the task demands? While a long research tradition suggests that L1 learners use both linguistic information and social information to resolve the indeterminacy of the referent problem (for example, Hollich et al., 2000), this possibility in L2 language acquisition has rarely been explored. In previous studies (Verga & Kotz, in prep.; Verga, Bigand & Kotz, in prep.) we have shown that learners may use both a knowledgeable partner as well as the sentence context to identify a new word referent; more specifically, a social partner may be particularly useful when the learner does not know a priori where the target referent may be. This result suggests that, indeed, the role played by a partner may be to drive the learner's visual attention towards a target, which, in turn, reduces uncertainty and facilitates learning.

In the current study, we used fMRI to investigate word learning in healthy adults during an online simulated social interaction. To this end, we employed a modified version of a contextual word learning game that we had validated and exploited in a previous series of studies (Verga et al., in prep.; Verga & Kotz, in prep.). In this set-up, participants performed an interactive game in which new words were repeated either in the same or in a different sentence context, displayed as a set of pictures presented on a checkerboard. During a training session, participants were divided into two groups: The first group (social group) performed the learning game with an experimenter, while the second group (non-social group) performed the task with a computer. Unbeknownst to the participants, during the learning session (which took place in the fMRI scanner) the game was controlled by a computer program irrespective of the group participants were assigned to. Consequently, all participants underwent the same experimental procedure, the only difference being that participants in the "social" group believed that they were interacting with a human partner. Since there was, in fact, no difference between participants trained "socially" and "non-socially" during the scanning procedure, this paradigm allows the impact of social interaction on contextual word learning to be tested (Mestres-Missé, 2007; Mestres-Missé et al., 2008, 2007, 2010; Rodríguez-Fornells et al., 2009). We explored contextual word learning in healthy adults using a 2-by-2 design with two factors – social interaction (social group, non-social group) and sentence context (same sentence context, different sentence context), which allowed us to test the relative impact of social and contextual cues on word learning. Further, the set-up allowed us to test specific hypotheses concerning three different moments in the task progression: First, the presentation of the stimulus, corresponding to the moment the learner is presented with all possible referents; second, the building-up of context, in which a social partner guides the learner towards the correct referent; and third, the word-meaning mapping, corresponding to the moment when the participant associates a novel word with its referent. We expected social interaction and sentence context to have a differential impact at the three different time points in learning.

First, we expected to observe an influence of social interaction on word learning, with an increase in activation of task-related areas for the social compared to the non-social group. In particular, we hypothesized that the presence of a social partner may help the learner by directing her/his attention towards the correct target referent (Dominey & Dodane, 2004; Tomasello, 2000; Verga & Kotz, 2013), leading to enhanced activity of the right inferior parietal cortex when the sentence context appears in the social compared to the non-social group (Carter & Huettel, 2013; Decety & Lamm, 2007). Since increased attention facilitates the processing of stimuli at attended locations (Bressler et al., 2008; Thiel et al., 2004), we expected social interaction to increase activations of areas involved in semantic decoding (encompassing IFG, MTG, MTL and subcortical structures including thalamus and striatum; Rodríguez-Fornells et al., 2009), since targets are, in this particular task, pictures from which the participant has to extract a meaning. We did not expect differences between social and non-social groups during the word-meaning mapping; indeed, this phase represented an individual moment, in which no help would be needed from the experimenter, as the participant simply had to silently read and memorize the new word.

We hypothesized that we would observe significant differences in activations corresponding to the processing of different types of sentence context. Indeed, as the learning phase proceeded, a sense of familiarity and ultimately recognition of a sentence would arise for items repeatedly presented in the same context (recruiting extra striate areas involved in object recognition and spatial navigation, and the precuneus, TPJ, and ACC; Huberle & Karnath, 2006), indicated behaviorally by faster reaction times and increased accuracy rates. Words repeated in a different context would instead lead to increased activtaion of areas involved in semantic retrieval, including the left inferior prefrontal cortex, left middle temporal gyrus, the anterior temporal lobes and subcortical structures (thalamus and striatum; Jeong et al., 2010; Mestres-Missé et al., 2008).

Resolution of the interaction between the two factors (social interaction and sentence context) will enable us to answer a crucial question – whether the use of social information that adult learners rely on depends on the task demands. As pointed out earlier and confirmed in our previous studies, it may be the case that adult learners use social information only if they really need it, and that they do not rely on a social partner if they can perform the task alone. This hypothesis supports an attention-driving account of social interaction (Dominey & Dodane, 2004; Tomasello, 2000; Verga & Kotz, 2013): When the participant does not know where to find a word referent (that is, in the different context condition), a social partner facilitates the task by directing her/his attention. Accordingly, we hypothesized that we would observe activations in areas critically involved in the re-orienting of attention, particularly the right angular gyrus (Carter & Huettel, 2013), as a result of the interaction between sentence context and social interaction in the different compared to the same context condition.

By itself, an increase in activation in a specific area would not justify the conclusion that a social partner directs a learner's attention. To prove this, it is necessary to consider how the network[21] supporting visuo-spatial attention is modulated by social interaction. To this end, we employed two statistical approaches, namely Independent Component Analysis and Dynamic Causal Modelling. ICA is an exploratory method, used to analyze fMRI images, which allows spatially independent signal components to be separated independently of the model assumed for the data (Calhoun, Adali, McGinty, et al., 2001; Calhoun, Adali, Pearlson, et al., 2001). For this reason, ICA is particularly useful as a control for standard fMRI analysis (Calhoun, Adali, McGinty, et al., 2001), especially when investigating processes for which a very clear model of the data is not yet available. Moreover, as a data driven approach it is especially powerful for confirming true positives in case of a priori hypotheses. Based on the literature reviewed above, we predicted that the social group compared to the non-social group would differ with regards to a correlation with the visuo-spatial attention network (Carter & Huettel, 2013; Dominey & Dodane, 2004;

---

[21] As pointed out by Calhoun and Adali (2012), 'network' is a somewhat ambiguous term that may be defined as a temporally correlated set of regions. From this perspective, a component identified with ICA defines a network, while singular foci of activation do not necessarily represent a network.

Tomasello, 2000; Verga & Kotz, 2013). Further, as the effect of attention should be a facilitation effect in the task-relevant visual cortices (via a top-down connection from the angular gyrus and superior visual cortices to the striate cortex; Chambers et al., 2004; Horwitz et al., 1998), we hypothesized that a stronger correlation would be found in the social group than in the non-social group, in areas involved in the visual analysis of the stimuli (in other words, the checkerboards and the elements of the sentence context).

If social interaction facilitates the learner by re-directing her/his attention towards the correct referent (Carter & Huettel, 2013; Dominey & Dodane, 2004; Tomasello, 2000; Verga & Kotz, 2013), and if the right parietal cortex works as a modulator of visual areas (Carter & Huettel, 2013; Elman et al., 2013), then the functional connectivity of areas involved in visuo-spatial attention should also be modulated by social interaction. In support of this hypothesis, it has been shown that persons suffering from Autistic Spectrum Disorders (ASD) have altered connections in the attentional network, with attention failing to modulate connectivity between extra striate areas and V1 during a visual task (Bird, Catmur, Silani, Frith, & Frith, 2006). We hypothesized that connectivity from the right angular gyrus to V1 would be significantly increased in the social compared to the non-social group. Further, we expected this modulation to occur when words were repeated in a different context, but not when they were repeated in the same context (when social information was most likely not of great use for the participant); indeed, as the learner will have been familiarized with the same context before, she/he will already know the correct target and consequently the re-orienting function mediated by right parietal cortex will not be necessary (Carter, Bowling, Reeck, & Huettel, 2012; Carter & Huettel, 2013). Instead, when the context changes each time a word occurs, the participant will not know where the referent will be; in this case, the help of the experimenter will be maximally useful, resulting in the engagement of the parietal lobe. We therefore employed DCM with a specific hypothesis-driven approach, which has the critical advantage of being able to specify directionality in the hypothesized effects of a condition on a network.

4.5.2. Methods

4.5.2.1 Participants

45 participants took part in the experiment. 4 participants were removed from the original sample because of excessive movement during scanning (exceeding the voxel dimension of 3 mm in either of the three directions). 41 participants (20F, mean age 26.98 years, SD 3.35) were included in the final data analyses. These were all native speakers of German recruited from the Max-Planck Institute for Human Cognitive and Brain Sciences database (Leipzig, Germany). All participants reported normal or corrected-to-normal vision, and none of them reported a history of hearing or neurological disorders. Right-handedness was assessed by the Edinburgh Handedness Inventory (Oldfield, 1971). All participants gave written informed consent and were paid for their participation. The experiment was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the University of Leipzig.

4.5.2.2 Material

The stimuli consisted of 180 checkerboards (3 x 3 pictures) each containing 9 images (330 x 245 pixels, 72 dpi) each centered in a different cell of the checkerboard (**Figure 8**). The images were black drawings presented on a light gray background to avoid excessive luminance contrast inside the scanner. They represented objects, humans, animals or actions and were originally selected from a validated database of pictures available online (Bates et al., 2003; Szekely et al., 2003, 2004, 2005; http://crl.ucsd.edu/experiments/ipnp/). A total of 49 images, all depicted in a singular form, were employed, including 12 pictures representing humans or animals (category: Subject), 17 representing actions (category: Verb), and 20 representing objects, humans or animals (category: Object). In each checkerboard, two nouns and an action were combined to form simple transitive German sentences (noun - transitive verb - target object, for example, “Der Junge isst das Ei”, “The boy eats the egg”). We defined “sentence context” as the combination of the subject and the verb. The pictures representing the elements of a sentence were arranged in such a way that the cells touched each other at least corner to corner. This constrained the game in the following ways: (i) Only one plausible sentence could be created within each checkerboard, and (ii) only one object could be chosen to form a plausible German sentence given a sentence context. The other six pictures in each checkerboard were distractor images chosen from the initial image pool, and balanced between pictures representing nouns (either animals, humans or objects) and actions. None of these distractor images constituted an additional plausible object in a given sentence context. The checkerboards were further balanced in terms of the mean naming frequency of the depicted items, and the mean number of times each element of the target sentence (subject, verb or object) appeared in each cell ($M_{subj} = 20.00$, $SD_{subj} = 3.57$; $M_{verb} = 20.00$, $SD_{verb} = 4.36$; $M_{obj} = 20.00$, $SD_{obj} = 2.83$). All possible dispositions (N = 69) for the three target images were employed a comparable number of times (M = 1.37, SD = 0.55).

The images belonging to the category “objects” (N = 20) were employed as targets for the sentence contexts. Each target was associated with a different Italian pseudo-word (length range: min. 4, max. 6 letters) extracted from a set of disyllabic pseudo-words (Kotz et al., 2010). The selected sample of pseudo-words was balanced for syllabic complexity, initial letter and final letter (“a” or “o”). We excluded words ending in “e” or “i” to avoid a possible confound with the Italian plural form, since all the pictures contained singular elements. Each pseudo-word and the associated target object could be presented a maximum of 9 times.

We manipulated 2 factors: 2 levels of social interaction (social interaction – S+, and non- social interaction – S-), and 2 levels of sentence context (different sentence context – dSC, and same sentence context – sSC).

To evaluate the impact of a social partner on the learning process, participants were randomly assigned to one of two conditions: Social interaction (S+, 10F, mean age 25.86 years, SD 2.87) or non-social interaction (S-, 9F, mean age 28.00 years, SD 3.77). Participants' age was balanced across the two groups [U = 136.50, z = -1.931, p = .054].

To evaluate the effect of sentence context variability, half of the objects (N = 10) occurred repetitively within the same sentence context (sSC – same sentence context). For example, the image representing "the cow" was always the correct ending for the same sentence context "the wolf bites". The other half of the objects (N = 10) was presented at each repetition within a different sentence context (dSC – different sentence context). For example, the image representing "the egg" could follow in sentence contexts such as "the woman cuts", "the boy eats", etc. Although each sentence was repeated 9 times, the actual number of exposures to each pseudo-word was dependent on the number of correct responses given by the participants, as a pseudo-word was presented only in case of correct responses. The alternation between sSC and dSC checkerboards was randomized.

4.5.2.3 Task and Experimental Procedure

4.5.2.3.1 Training phase

Participants received detailed written instructions and performed a block of 10 practice trials to familiarize themselves with the task requirements. This phase was additionally used to establish a distinction between a socially interactive condition and a non-social condition. Upon their arrival in the lab, participants were pseudo-randomly assigned to either of the two conditions. Participants assigned to the “social” group were told they would do the experiment with an experimenter who was positioned in front of the controlling computer outside the scanner room. To enforce this belief, these participants performed the practice trials in a truly interactive fashion with the experimenter, following the procedure we established in previous studies (Verga et al., in prep.; Verga & Kotz, in prep.).

The participants’ task was to find the correct ending for a given sentence context from amongst the checkerboard images. In all conditions, each trial began with the presentation of a fixation cross (500 ms), followed by a checkerboard. In each checkerboard, the sentence context was provided to the participant, signaled by a red frame appearing in succession around the subject and verb of the sentence context. When both elements were marked with a red frame, participants could select from among the remaining 7 images in the checkerboard an object fitting the sentence context. The experimenter selected the subject and verb of the target sentence for participants in the social condition, while a computer program did the same for participants assigned to the non-social condition. In this phase there was no time limit to answer; however, participants were informed that during the learning phase they had up to 4 seconds to provide their answer. In both groups, when a correct answer was given the selected image was substituted by a pseudo-word providing the “Italian name” of the object. Pseudo-words were presented in black capital letters over a light grey background in the selected object cell, and remained on the screen for 1000 ms (font Arial, size 40 points). If an incorrect response was given, no “Italian name” was displayed and the following trial began immediately.

4.5.2.3.2 Learning Phase

After this training, participants performed the learning phase during which fMRI brain images were acquired. The learning phase was identical to the training phase. Participants belonging to the “social” group were told that the experimenter was controlling the program from outside the scanner room, whereas participants in the non-social group were told that they were engaged in a computer task. However, there was no real difference between these two conditions, as in both cases the stimulus presentation was controlled by the same computer program without the experimenter's intervention. In total, this phase had a duration of ~48 minutes and was divided into 3 functional runs lasting 16 minutes each. 180 stimuli (60 x run) were presented, alternating with 60 null events (20 x run), and each run contained 3 repetitions for each target object (9 total). The selection of the subject picture was jittered (range: 2.63 – 4.65 sec, M = 3.59 sec, SD = 0.56 sec) to ensure a better sampling of the haemodynamic response. The selection of the verb picture followed the subject picture with a delay of 500 ms, to mimic the real behavior of a human experimenter, as seen in our previous studies (Verga et al., in prep.; Verga & Kotz, in prep.)**.**

In both groups (S+ and S-), the experimenter communicated with participants during the pauses between functional runs. The experimenter asked participants of the social group whether her speed in providing the sentence context was adequate, or if a different pace was required. This ploy was used to strengthen the participants' belief that the experimenter was interacting with them. For participants belonging to the non-social group, the experimenter simply asked if everything was proceeding smoothly. Responses were collected using an fmri-compatible, in-house touch pad, allowing the participant to select the correct object on the checkerboard. The touch-pad was positioned under the participants' right hand. To confirm her/his choice, participants had to press a response-button with their left thumb. Participants were carefully instructed not to move too much on the touch pad and to only move the finger to select the identified target. Visual stimuli were presented via back-projection (projector SANYO PLC-XP50L) onto a screen positioned at the rear-end of the bore, and were visible by the participants via a mirror mounted on the head-coil.

4.5.2.3.3 Testing phase

At the end of the learning phase, a behavioral testing phase took place outside the scanner room to evaluate whether pseudo-words presented during the learning phase had been correctly mapped to the corresponding objects. In this task, participants were presented with combinations of subject and verb pictures that had not been seen together before, followed by three of the "Italian words" participants learned during the learning phase. The participants' task was to select the "Italian word" indicating an object that fitted the given sentence context. All trials contained one correct and two incorrect options.

Lastly, participants were presented with a paper-and-pencil questionnaire to evaluate the extent to which social simulation had been effective. The questionnaire consisted of 12 questions primarily intended to verify whether "social" participants believed the cover story, and included indirect questions concerning the experimenter's performance (for example, "Was your partner in the game too slow?"), as well as more direct questions (for example, "Were you sure you partner in the game was a real person?"). Participants in the non-social condition filled in a similar 12-question questionnaire in which "the partner" was substituted with "the computer". A copy of the questionnaire is presented in Appendices A (social questionnaire) and B (non-social questionnaire).

All stimuli were presented using a desktop computer running Presentation 16.0 (Neurobehavioral Systems, Albany, USA). For the practice trials in the social condition, two standard wheel mice (Logitech Premium Optical Wheel Mouse) were connected to the same Windows laptop and used as response devices. For the practice trials in the non-social condition, and the testing phase in both groups, only one standard wheel mouse (Logitech Premium Optical Wheel Mouse) was employed. At the end of the experiment, participants were debriefed about the real set-up and aim of the experiment. A new participant was admitted to the training room only after the previous one left, to avoid him/her hearing the debriefing and revealing the cover story. All participants in the S+ condition confirmed that they believed the cover story.

4.5.2.4 fMRI data acquisition

Brain images were acquired using a 3T whole-body Siemens TRIO scanner (Siemens Healthcare, Erlangen, Germany) at the Max Planck Institute in Leipzig. Functional whole-brain images were collected using a T2*-weighted gradient echo, echo-planar (EPI) pulse sequence, using the Blood Oxygenation Level Dependent (BOLD) contrast. Each functional volume consisted of 30 axial slices parallel to the AC-PC line with a slice thickness of 3 mm and a 33% inter-slice gap (TR = 2000 ms, TE = 30 ms, flip angle 90°, FOV 192.2 cm, matrix size 64 x 64, acquisition bandwidth 116 kHz). The first 2 volumes in each run were discarded to account for saturation effects.

A field map (short TE = 4.92 ms, long TE = 7.38 ms, echo spacing = 0.69 ms) was acquired for each participant at the beginning of the session to provide an estimate of local field inhomogeneity and later used in the spatial pre-processing of functional images.

Previously acquired high-resolution anatomical images were used for co-registration of the functional data. These images were T1 weighted 3D MP-RAGE structural scans with a sagittal orientation and a spatial resolution of 1 x 1 x 1.5 mm (TR = 1300 ms, TE = 3.93 ms, FOV 256 x 240 mm, acquisition bandwidth 67 kHz).

4.5.2.5 fMRI data analyses

4.5.2.5.1 Pre-processing

Data pre-processing and statistical analysis were conducted using SPM8 (Statistical Parametric Mapping, Wellcome Trust Centre for NeuroImaging, London, UK) running on MATLAB 7.11 version (The Mathworks Inc., Natick, USA). For each participant, slices were corrected for differences in acquisition time, realigned to the first image of the first session and unwarped to correct for field inhomogeneity. Images were then normalized to the Montreal Neurological Institute (MNI) standard space and smoothed with an 8 mm full-width at half-maximum Gaussian isotropic kernel.

4.5.2.5.2 Standard GLM Analysis

Data were statistically analyzed with a standard general linear model procedure in two stages. At the first level, regressors for the experimental conditions were convolved with a canonical haemodynamic response function. The design matrix included single regressors for all the variables of interest. Our trials contained 3 different moments with specific characteristics, and involved different learning aspects: First, the appearance of the checkerboards; second, the identification of subject and verb; third, the appearance of the pseudo-word when a correct response was given (**Figure 25**). These three moments (checkerboard, sentence context, object) were explicitly modeled in the design matrix for each sentence context condition. Further, to model possible differences in difficulty between conditions (sSC and dSC), participants' response times were added to the design matrix as parametric modulations of object sSC and object dSC regressors. Error trials were modeled as an additional regressor of no interest. To evaluate the specific effect of learning, repetitions for each trial were added to the design matrix as parametric modulators of the respective experimental event and functional run (1, 2 and 3) were included as constant regressors. Temporal series for each subject were filtered with a 100 Hz high-pass filter to remove slow signal drifts, and first-level Student's t-tests for each experimental condition were calculated (checkerboard [sSC], checkerboard [dSC], subject [sSC], subject [dSC], object [sSC], object [dSC]). All the events were considered events of null duration, except the object encoding that was considered as an epoch of 2 seconds (that is, the duration of the word presentation, corresponding to a passive stimulation). Contrasts directly comparing the single conditions were also calculated at the first level (for example, sSC > dSC). At the second level, t-contrasts defined at the first level were fed into one-sample (within-subject comparisons) and paired-sample (between-subject comparisons) t-tests.

Correlations between brain activity and behavioral scores (accuracy scores during the testing phase) were estimated with a multiple regression analysis at the second level. All results for the within-subject comparisons are reported as FWE-corrected at a p-level of .05 and a cluster extension of $k > 20$, unless otherwise specified. As between-group comparisons increase the variance in the sample, and consequently reduce statistical power, a more liberal criterion has been used in these contrasts to control for the probability of obtaining false positive results. To this end, we employed a cluster extent threshold criterion (Forman et al., 1995). Monte Carlo simulations were conducted using the program Alphasim implemented in AFNI software; the threshold for single voxels was set at $p < .001$ and a Monte Carlo simulation with 2000 iterations was run, resulting in an extent threshold of 30 voxels to achieve a correct threshold of $p < .05$.

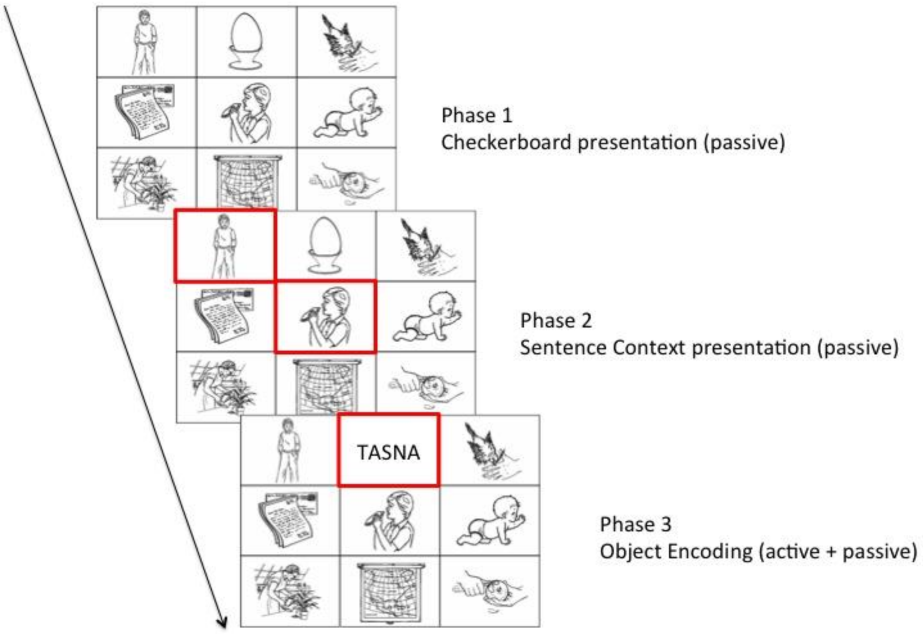
Images are displayed superimposed onto the skull-stripped mean anatomical image of the 41 subjects, and warped to the MNI space. Tables containing activation coordinates and significance values are reported in Appendix E.

4.5.2.5.3 Independent Component Analysis (ICA)

An Independent Component Analysis was conducted using the ICA group analysis toolbox for fMRI data (GIFT, Calhoun et al., 2001), implemented in SPM5 (Statistical Parametric Mapping, Wellcome Department of Imaging Neuroscience, London, UK). The motion-corrected, normalized and smoothed images from each session and each participant were reduced by means of a

Figure 25 - Learning phases example.

The illustration depicts the three moments of the task modeled in the analysis. First, the presentation of the checkerboards, which corresponds to the moment the learner is presented with a new environment containing all the possible referents. Second, the presentation of the sentence context, which corresponds to the moment when the experimenter directs the learner's attention towards the referents. Third, the encoding of the object, when the correct referent has been identified and the word-meaning mapping takes place.



PCA, in two steps (30 principal components in the first steps were further reduced to 20 in the second step). The ICA analysis was subsequently carried out using the Infomax ICA algorithm (Bell & Sejnowski, 1995). Group analysis was performed with the ICASSO procedure and each individual participant's data were then back reconstructed with the GICA3 method and scaled to z-scores. One-sample t-tests were performed to evaluate the significance of independent component maps. Results were accepted as significant with a $p < .05$ voxel level, FWE corrected. The ICA time course for each component, experimental condition, and participant were temporally sorted and linearly regressed with the design matrix stimulus onset parameters for each experimental condition. All results are reported as FWE-corrected at a p-level of .05. The resulting beta estimates of this first-level multivariate regression were entered into a group-level, random-effects, one-sample t-test analysis to estimate the size of the positive or negative correlation between the stimulation for each experimental condition and the activation time course of a particular independent component.

4.5.2.5.4 Dynamic Causal Modelling (DCM)

DCM is a method used to investigate effective connectivity, and to infer hidden neuronal states from measurements of brain activity (Friston et al., 2003). While effective connectivity is often contrasted with functional connectivity, both families of methods in fact use measures based on the BOLD signal; as such, they are different from other measures investigating structural connectivity through white matter assessment (Mori & Zhang, 2006)[22]. The difference between functional and effective connectivity methods rests in the possibility of specifying causal relations between distal regions; while

---

[22] It must be pointed out that the distinction between structural and functional/effective measures of connectivity is mostly methodological. Indeed, while based on different measures, structural and functional/effective connectivity are not independent and reflect two faces of the same coin (Kahan & Foltynie, 2013; Stephan, Tittgemeyer, Knösche, Moran, & Friston, 2009).

functional connectivity methods only describe correlations between brain activity in different areas, effective connectivity methods consider the direct influence that a region exerts on another (Kahan & Foltynie, 2013; Stephan et al., 2010).

Based on our GLM results and on our a priori hypotheses supported by the literature, we used DCM (Friston et al., 2003) to test the hypothesis that the fronto-parietal attention network involved in attentional re-orienting would be activated more in the social compared to the non-social group. More specifically, we expect to see higher connectivity in the social group compared to the non-social group for the backward connections from the right angular gyrus to right V1 (Decety & Lamm, 2007).

In order to investigate this question, we specified two dynamic causal models, one for each condition (sSC and dSC), including three regions in the right hemisphere (rMFG, rAG and rV1). We fitted two identical but separate models per condition (sSC and dSC) in order to allow all parameters to possibly change between the two conditions. In this analysis, we grouped together the trial event corresponding to the appearance of the checkerboard and sentence context, as both events require a re-direction of attention towards the novel visual stimulus. The two models were created for each participant on a GLM model including 2 regressors for the two conditions (dSC and sSC) for each of the three sessions and three constant regressors. These first-level GLMs were pre-processed and analyzed in this first step as described in paragraph 4.5.2.5.2.

The three regions of interest (rMFG, rAG and rV1) were defined as 6mm-radius spheres centered for each subject and each session on the maxima of statistic parametric maps, testing for the overall main effect of different context (dSC) or same context (sSC) on the network. Regional activations were extracted in terms of the principal eigenvariate from each region, in a run-specific fashion. The following MNI coordinates were chosen for the three volumes of interest (VOIs) that constitute our model: rV1: x = 18, y = -85, z =

8; rAG: x = 48, y = -64, z = 43; rMFG: x = 33, y = 14, z = 61. We limited our research to the right hemisphere since our main focus was on the spatial-attention network, consistently reported to be right lateralized (Chambers et al., 2004). The volumes of interest were corrected for the effects of interest (omnibus F-test), so that they were not biased towards any particular experimental conditions, but instead included the information relative to the stimuli of all conditions.

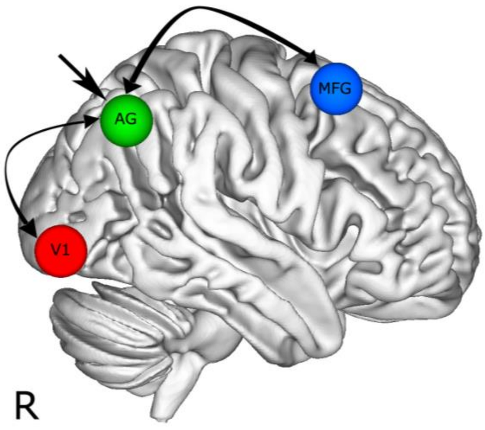
Our model is based on previous evidence that the parietal cortex supports spatial attention by working as a junction between the middle frontal gyrus (BA 8, Han et al., 2004; Verghese, Kolbe, Anderson, Egan, & Vidyasagar, 2014) and the visual cortex (Constantinidis et al., 2013; Macaluso et al., 2000; Saalmann et al., 2007), with the function of re-directing attention towards new targets (Corbetta et al., 2000, 2008). More specifically, superior areas such as middle frontal gyrus (Thiel et al., 2004) facilitate the activity of the task-relevant visual cortices via top-down connections from the angular gyrus and superior visual cortices to the striate cortex (Chambers et al., 2004; Horwitz et al., 1998), ultimately facilitating the processing of stimuli at attended locations (Bressler et al., 2008; Thiel et al., 2004). Accordingly, we defined our model (for both conditions) as a fully connected model with reciprocal intrinsic connections from V1 to the angular gyrus (v1 ↔ rAG) and from the angular gyrus to and from the middle frontal gyrus (rAG ↔ rMFG) in the right hemisphere. A graphical representation of the model is presented in **Figure 26**.

Intrinsic connectivity analysis was performed using the DCM10 tool supported by SPM8. For each connection specified in the model, we computed the coupling parameters across functional runs and used one-sample t-tests (IBM SPSS Statistics 18, IBM Corporation, New York, USA) to ensure that each connection was significantly different from 0. Since the current study aimed to explore the learning process under social conditions, we computed mean parameters for the regional connectivity established by the two experimental conditions for each functional run (run1, run2 and run3). We then used a repeated measures ANOVA (two within-subject factors: run – run1, run2, run3, and sentence context – sSC, dSC; one between-subject factor: social interaction – S+, S-) to assess the difference in connection strengths across runs and between groups, and a possible interaction of these two factors. We performed a second-level random-effects analysis on the connection parameters outside SPM using IBM SPSS Statistics 18 (IBM Corporation, New York, USA).

For both the intrinsic and modulatory connection strengths, we first used a Shapiro-Wilk normality test to check the normality of the distribution of values pertaining to each connection. To control for the number of tested connections (4 for each model) we calculated alpha values corrected according to the false discovery rate (FDR; Benjamini & Hochberg, 1995).

Figure 26 - Representation of the DCM model.

Three VOIs were selected corresponding to the visual area V1, the angular ayrus (AG) and the middle frontal gyrus (MFG) in the right hemisphere. The input is shown entering the model in the angular gyrus.



4.5.2.6 Behavioral data analysis

Statistical analyses of behavioral data were performed using MATLAB R2013a (The Mathworks Inc., Natick, USA) and IBM SPSS Statistics 18 (IBM Corporation, New York, USA).

Behavioral data were first corrected for outliers; trials with response times exceeding the mean response times (RTs) ± 2 standard deviations (SDs) were excluded from further analysis (run1: $M_{rt} = 3.99$, $SD_{rt} = 0.66$; run2: $M_{rt} = 3.18$, $SD_{rt} = 0.82$; run3: $M_{rt} = 2.86$, $SD_{rt} = 0.94$; mean of rejected trials across participants = 6.05 %).

For the training and learning phases, response times were calculated as the time delay between the appearance of the “verb” image and the participant’s answer. Accuracy scores (proportion of correct responses), response times of correct responses and their standard deviations were calculated at each repetition of the object for each participant. In previous studies, we have shown that social interaction has an impact on the degree of temporal coordination between the experimenter and the participant; this increase in temporal coordination is consistently reported in the literature as a marker of social interaction (Yun et al., 2012). For this reason, even in this study we calculated indexes of temporal coordination. First, SDs of response times were employed as an index of stability of participants’ performance; in other words, the higher the SDs, the less stable (or more variable) the performance. Further, we calculated the lag-0 and lag-1 cross correlation coefficients between the inter-trial-intervals produced by the participants, and those produced by the experimenter (S+ condition) or the computer (S- condition). These measures indicate the relation between the temporal behavior of the experimenter/computer and the participant. More specifically, the Cross Correlation (cc) at lag-0 indicates how much the behavior of the participant in one trial is temporally related to the behavior of the partner (experimenter/computer) in the same trial; in other words, they indicate online temporal coordination. The Cross-Correlations at lag-1 indicate whether the behavior of the experimenter/computer was related to the participant's behavior in the following trial.

To account for the difference in variability of trial presentation in the S+ compared to the S- condition in the training phase, we conducted separate ANCOVAs on the variables of interest. We used SDs of the experimenter's response/computer times as covariates during the learning phase, in all analyses except the cross-correlation; indeed, in the latter case SDs account for the variability in the computer/experimenter RTs series, on which correlation coefficients are calculated.

Response times from the testing phase were calculated as the time delay between the appearance of the pseudo-words and the participant's response, and accuracy scores were defined as the proportion of correct responses. We reasoned that the number of exposures during the learning phase could have an impact on the responses' accuracy and response times; in other words, participants who saw the association picture/pseudo-word a greater number of times would memorize the word better. We therefore used the mean number of repetitions to which the participants were exposed to as a covariate in the analyses. This number is directly related to the number of correct responses and takes into account the mean number of times that pictures were repeated during the learning phase, ranging from a minimum of 0 (no correct responses) to a maximum of 9 times (no errors).

When the assumption of sphericity was not met, a Greenhouse-Geisser correction was applied to the degrees of freedom. Two-tailed t-tests and simple-effect analyses were employed to compare individual experimental conditions. We used an alpha level of $p < .05$ to ascertain significance for all statistical tests, and a Bonferroni correction was applied when required, to account for multiple comparisons.

4.5.3. Behavioral results

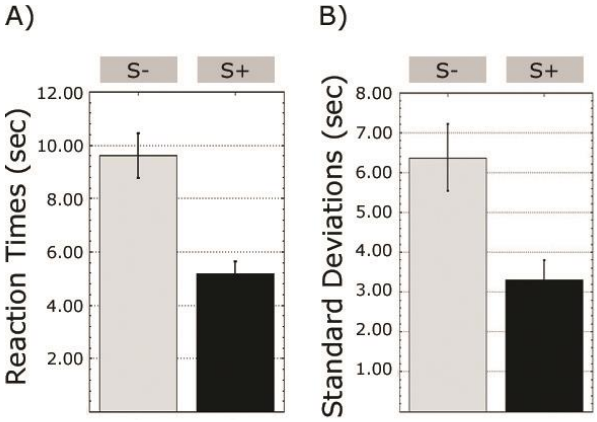
4.5.3.1 Training phase

Separate analyses of variance (ANCOVA) were conducted on response times, accuracy scores, standard deviations and cross-correlation indexes, while controlling for differences in variability between experimenter and computer presentation. In this phase, we could only compare the social and non-social groups, as the sentence context variability factor was specified across repetitions and was not applicable in ten trials.

Participants reached an accuracy level of 82%. There was no difference in response ***accuracy*** between participants trained socially or non-socially [F(1,38) = .123, p = .728, η_p^2 = .003]. However, ***response times*** were significantly faster for participants doing the training with the experimenter [M

Figure 27 – Training phase.

A) Response times of correct responses during the training phase. B) Standard deviations of RTs during training. Vertical bars represent standard error of the mean. Abbreviations: S- = non-social interaction; S+ = social interaction; dSC = different sentence context; sSC = same sentence context.



= 4.447, SD = 2.137] compared to participants doing the task alone [M = 10.364, SD = 3.695] [F(1,38) = 20.080, p = .000, η_p^2 = .346] (**Figure 27A**). Further, ***standard deviations*** of response times were significantly smaller in the social group [M = 3.306, SD = 2.223] compared to the non-social group [M = 6.665, SD = 3.780] [F(1,38) = 8.256, p = .007, η_p^2 = .178] (**Figure 27B**). There was no difference between groups either for ***lag-0*** [F(1,39) = 0.087, p = .769, η_p^2 = .002], or for ***lag-1 cross-correlations*** [F(1,39) = .588, p = .448, η_p^2 = .015].

4.5.3.2 Learning phase

Participants' ***accuracy*** significantly increased across runs [linear trend, F(1.455, 56.731) = 152.644, p = .000, η_p^2 = .796]. Further, accuracy was higher for sSC (M = .798, SEM = .017) as compared to dSC (M = .693, SEM = .019) items [F(1,39) = 66.026, p = .000, η_p^2 = .629]. The interaction between these two factors was also significant, showing a progressively increasing gap

Figure 28 – Learning phase response accuracy (proportion of correct responses).

Abbreviations: S- = non-social interaction; S+ = social interaction; sSC = same sentence context; dSC = different sentence context.

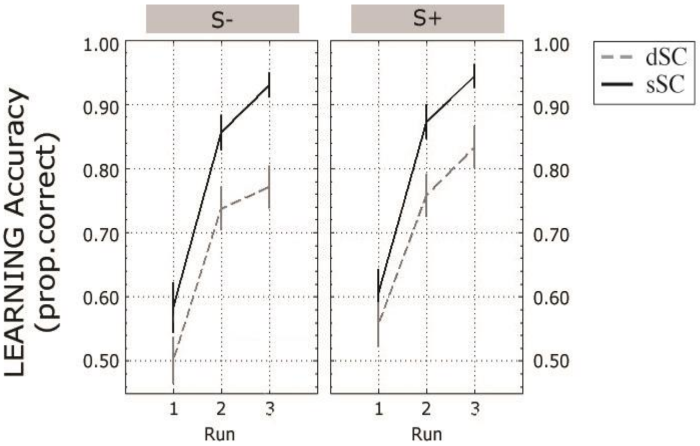
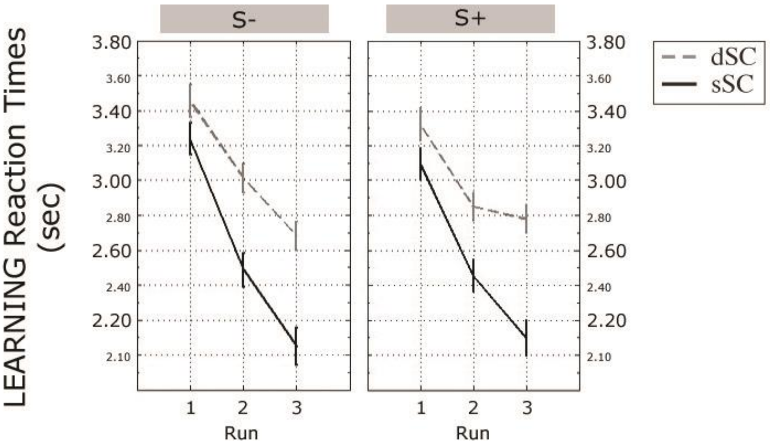


Figure 29 – Learning phase response times.

Abbreviations: S- = non-social interaction; S+ = social interaction; sSC = same sentence context; dSC = different sentence context.



between the two context conditions [F(1.677, 65.417) = 6.498, p = .000, η_p^2 = .143]. There was no effect of social interaction [F(1,39) = .905, p = .347, η_p^2 = .023] and no further interactions (all ps > .282; **Figure 28**).

The ANOVA conducted on ***response times*** showed a significant decrease from one functional run to the next [linear trend, F(1.575,61.438) = 180.668, p = .000, η_p^2 = .003]. Furthermore, sSC items elicited faster response times compared to dSC items [F(1,78) = 252.285, p = .000, η_p^2 = .866]. In addition, the interaction between run and sentence context was significant, showing increasingly faster RTs in the sSC condition compared to the dSC condition, especially in the second and third run [F(1.575,61.438) = 28.794, p = .000, η_p^2 = .425] (**Figure 29**). There was no effect of social interaction [F(1,39) = .240, p = .627, η_p^2 = .006] and no further interactions (all ps > .084).

There were no significant effects on the ***standard deviations*** for any of the dependent measures (all ps > .070).

C***ross-correlation coefficients at lag-0*** were not significantly different between sSC and dSC [F(1,38) = .094, p = .760, η_p^2 = .002] or across the runs [F(2,76) = .028, p = .972, η_p^2 = .001]. However, the interaction between sentence context and run was significant [F(2,76) = 7.923, p = .000, η_p^2 = .173], showing that sSC and dSC conditions were significantly different in the first and last run [run1: p = .002; run2: p = .037], but not in the second run (p = .563). There was no difference between social and non-social groups [F(1,38) = .422, p = .520, η_p^2 = .011]. There were no other effects or interactions (all ps > .692).

C***ross-correlation coefficients at lag-1*** were non-significant for all factors (all ps > .060).

4.5.3.3 Testing phase

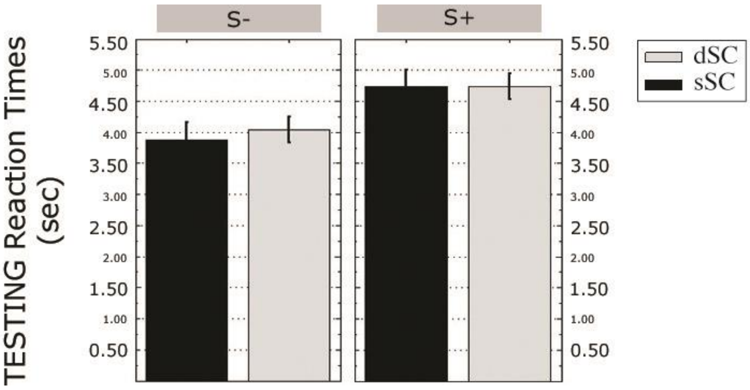
Separate analyses of covariance (ANCOVA) were conducted on accuracy scores and response times during the testing phase, while controlling for the mean number of item repetitions during the learning phase. Overall, participants performed at an average accuracy level of 70%.

The analysis of **response times** revealed that participants trained socially were significantly slower [M = 4.737, SD = 1.214] than participants trained non-socially [M = 3.960, SD = .839] [F(1,38) = 6.767, p = .013, η_p^2 = .151] (**Figure 30**). There was no effect of sentence context [F(1,38) = .034, p = .854, η_p^2 = .001] and no interaction between sentence context and social interaction [F(1,38) = .242, p = .626, η_p^2 = .006].

There was no difference in the **accuracy** scores between participants trained socially or non-socially [F(1,38) = 2.777, p = .104, η_p^2 = .068], or between items that were repeated consistently or more variably [F(1,38) = 1.184, p = .283, η_p^2 = .030], and no interaction between social interaction and sentence context [F(1,38) = .392, p = .535, η_p^2 = .010].

Figure 30 – Testing phase response times.

Abbreviations: S- = non-social interaction; S+ = social interaction; SC = same sentence context; DC = different sentence context.



4.5.3.4 Summary of behavioral results

During the initial training phase, participants trained socially were significantly faster and less variable than participants trained with a computer.

However, during the learning phase, no behavioral differences were observed between the two groups of participants. Words encoded in a more consistent sentence context were recognized faster, and with greater accuracy, than words repeated in a variable sentence context.

Lastly, in the testing phase, participants trained socially gave significantly slower responses compared to participants trained non-socially.

4.5.4. fMRI results, Part I: Whole-brain GLM analysis

4.5.4.1 Main effects of social interaction

Checkerboard Observation: At a conservative threshold of $p < .05$, FWE-corrected, there were no significant activations for either of the planned contrasts. However, significant activations at a less stringent threshold ($p < .001$, $k > 15$) were found for the contrast **checkerboard [S+ > S-]** in the **bilateral hippocampi**. These results are displayed in **Figure 31**, and **Table 16** (Appendix E).

Furthermore, at the liberal threshold of $p < .001$ (uncorrected), the correlation between lag-0 cross-correlations and activity in the left middle temporal gyrus (**Figure 32** and **Table 17**, Appendix E) was significantly higher for participants in the S+ group compared to participants in the S- group.

Subject Extraction: At the liberal threshold of $p < .001$, clusters of significant activations were found for the contrast **subject [S+ > S-]** in the left middle occipital gyrus and left inferior temporal gyrus. These results are displayed in **Figure 33** and **Table 18** (Appendix E).

Object Encoding: No main effects of social interaction were found in this phase, at any significance threshold.

Learning effects: The interaction of run and social interaction (contrast: **subject [(run3 > run1)*(S+ > S-)])** elicited a significant activation in the right supramarginal gyrus and in the cerebellum bilaterally (**Figure 34** and **Table 19,** FWE-corrected with a small volume correction from Jeong's coordinates, radius = 20 mm).

Figure 31 - Contrast: checkerboard [S+ > S-].

Activations are displayed on axial slices of the average anatomical image of all participants warped to the MNI system of coordinates (p < .001, k > 15).

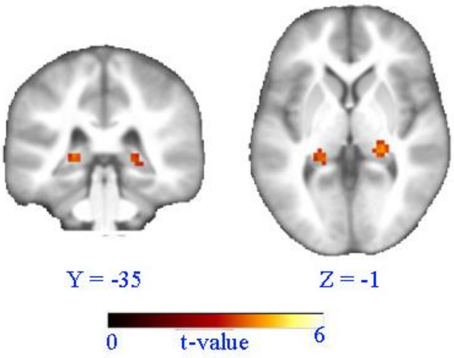


Figure 32 - Correlation: checkerboard [r (lag-0, (S- > S+))].

Areas displaying a different correlation between groups with the lag-0 cross-correlations are displayed (p < .001, k > 20).

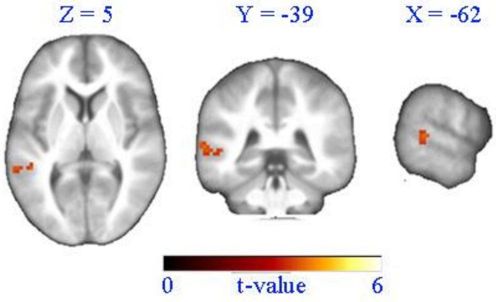


Figure 33 - Contrast: subject [S+ > S-].

Significant activations are displayed on axial slices of the average anatomical image of all participants warped to the MNI system of coordinates (p < .001, k > 12).

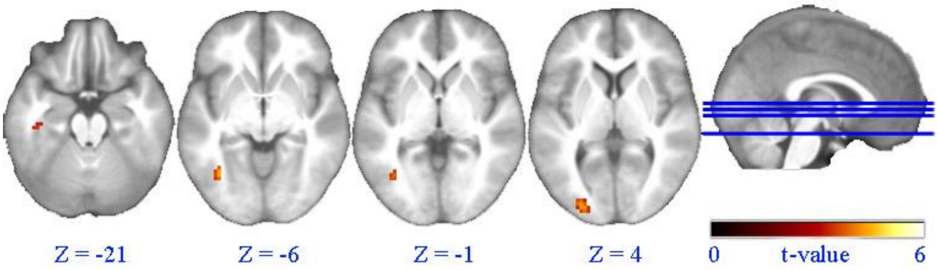
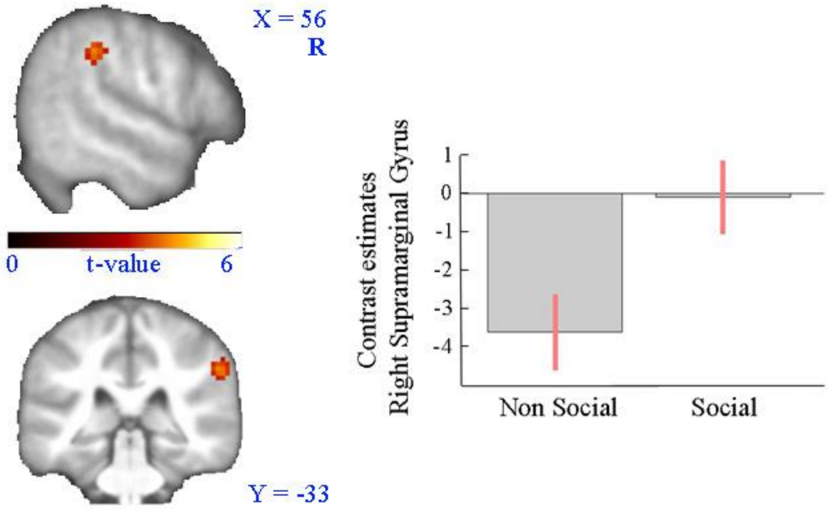


Figure 34 - Contrast: subject [(run3 > run1)*(S+ > S-)].

Significant activations are displayed on axial slices of the average anatomical image of all participants warped to the MNI system of coordinates (p < .001, k > 20).



4.5.4.2 Main effects of sentence context

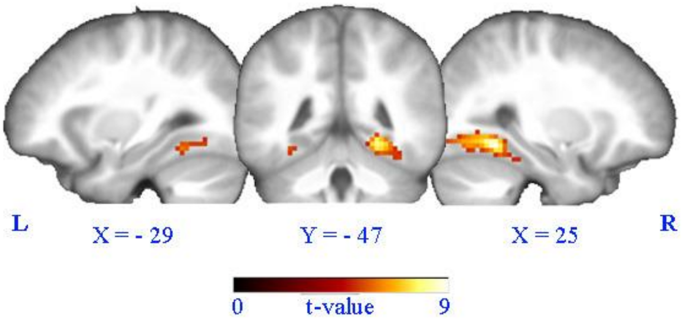
Checkerboard Observation: The contrast **checkerboard [dSC > sSC]** elicited significantly greater haemodynamic responses in the bilateral fusiform gyri and the right cerebellum. Results are summarized in **Figure 35** and **Table 20** (Appendix E).

Subject Extraction: The contrast **subject [sSC > dSC]** elicited activations in a bilateral set of areas encompassing the bilateral cerebellum, inferior parietal lobe and superior temporal gyrus. In the right hemisphere, the middle and posterior cingulate cortex and putamen were additionally activated. In the left hemisphere, significant activations were found in the amygdala and insula lobe (**Figure 36a** and **Table 21a**, Appendix E).

Results for the opposite contrast **subject [dSC > sSC]** are summarized in **Figure 36b** and **Table 21b** (Appendix E). Areas displaying significantly greater activations include the bilateral inferior frontal gyrus and fusiform gyrus. Further activations were observed in the left hemisphere, in the lingual and

Figure 35 - Contrast: checkerboard [dSC > sSC].

Activations are displayed on axial slices of the average anatomical image of all participants warped to the MNI system of coordinates (p < .05, FWE-corrected for multiple comparisons, k > 30).

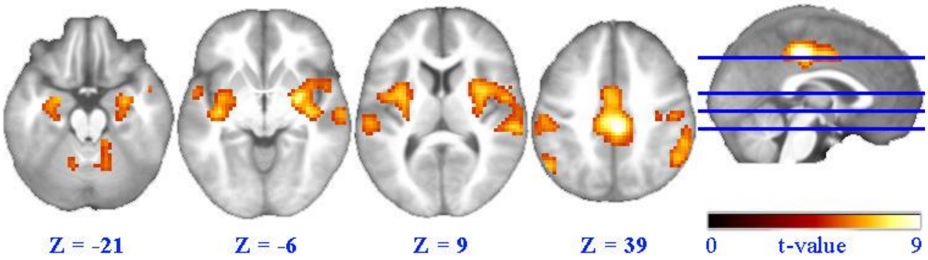


superior occipital gyri, while in the right hemisphere, the inferior and middle occipital gyrus and the inferior temporal gyrus were activated.

Object Encoding: The contrast **object [dSC > sSC]** elicited significant clusters of activations in the bilateral inferior frontal, middle temporal and pre-central gyri. In the right hemisphere, significant activations were found in the

Figure 36 - Subject Extraction: sentence context (p < .05, FWE corrected for multiple comparisons, k > 30)

36a) Contrast: subject [sSC > dSC]. Significant activations are displayed on axial slices of the average anatomical image of all participants warped to the MNI system of coordinates.



36b) Contrast: subject [dSC > sSC]. Significant activations are displayed on axial slices of the average anatomical image of all participants warped to the MNI system of coordinates.

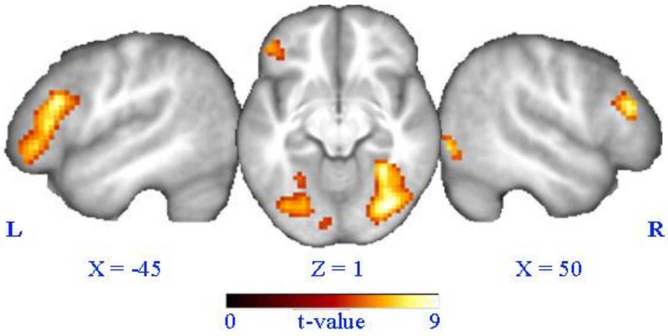
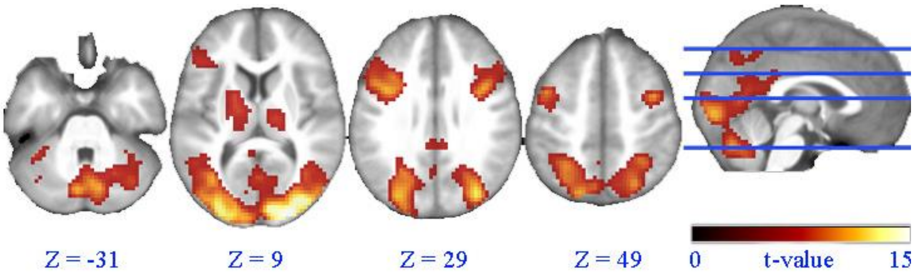


Figure 37 - Contrast: object [dSC > sSC].

Activations for the contrast object [dSC > sSC] are displayed on sagittal slices of the average anatomical image of all participants warped to the MNI system of coordinates (p < .05, FWE-corrected for multiple comparisons, k > 30).



middle frontal gyrus and the thalamus. In the left hemisphere, further activations were found in the inferior temporal and occipital gyrus. Results are summarized in **Figure 37** and **Table 22**.

4.5.4.3 Interactions between sentence context and social interaction

Checkerboard Observation: In the different context condition, the social group displayed a higher correlation between **test scores** and activity in the right inferior frontal gyrus, pars orbitalis (**Figure 38** and **Table 23**). Furthermore, activation in the right lingual gyrus was significantly more correlated with the **test scores** in the non-social compared to the social group (**Figure 39** and **Table 24**).

In the same context condition, the correlation between activity in the right caudate nucleus and lingual gyrus and the lag-0 cc index was significantly higher in the social compared to the non-social group (**Figure 40** and **Table 25**).

Subject Extraction: At the liberal threshold of p < .001 (uncorrected), the contrast **subject [(sSC > dSC) * (S+ > S-)]**, analyzing the interaction between sentence context and social interaction, elicited significant activations in the right angular gyrus and middle frontal gyrus (**Figure 41** and **Table 26**).

<u>Object Encoding:</u> At a threshold of p < .05 corrected (extent threshold criterion) we observed a cluster of activation in the left cerebellum (lobule V) as a result of the interaction between **social interaction and sentence context** (contrast: **object [(sSC > dSC) * (S+ > S-)]**. Results are displayed in **Figure 42** and **Table 27**.

Figure 38 - Correlation: checkerboard [test scores (dSC (S+ > S-))].

Areas displaying a positive correlation with the test scores during checkerboard observation (p < 0.05, corrected).

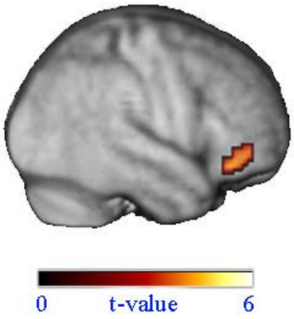


Figure 39 - Correlation: checkerboard [test scores (dSC (S- > S+)].

Areas displaying a positive correlation with the test scores during checkerboard observation ($p < 0.05$, corrected).

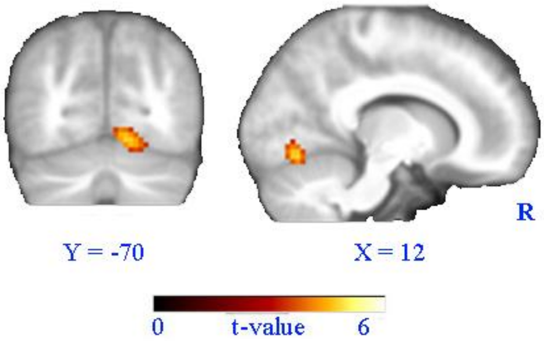


Figure 40 - Correlation: checkerboard [lag0 (sSC (S+ > S-))].

Areas displaying a correlation with lag-0 cross-correlations in the same context condition are displayed ($p < .05$, FWE-corrected for multiple comparisons at the cluster level, $k > 30$).

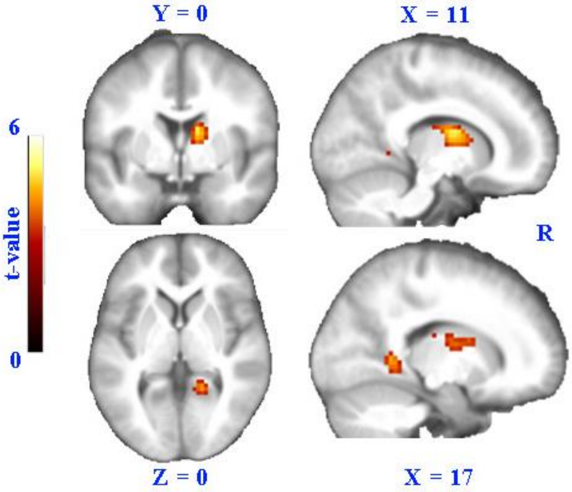


Figure 41 - Contrast: subject [(sSC > dSC)*(S+ > S-)].

Significant activations are displayed on axial slices of the average anatomical image of all participants warped to the MNI system of coordinates (p < .001 k > 13).

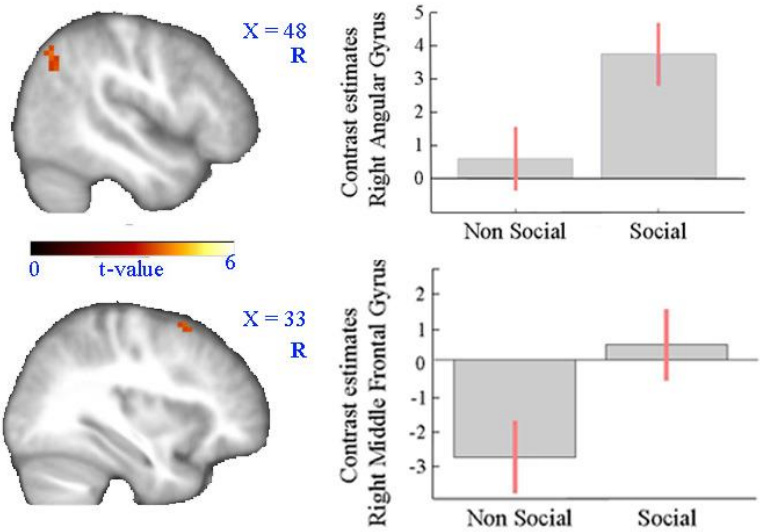
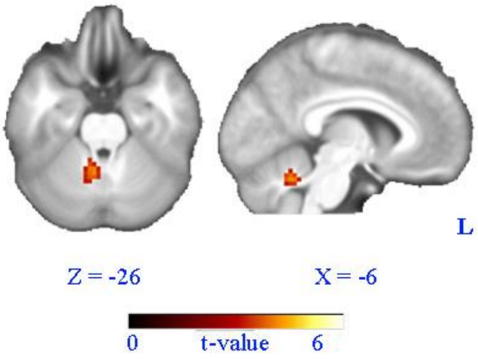


Figure 42 - Contrast: object [(S+ > S-)*(sSC > dSC)].

Activations are displayed on axial slices of the average anatomical image of all participants warped to the MNI system of coordinates (p < .05, k > 30).



4.5.5. fMRI results, Part III: ICA

20 components were evaluated with ICA and subsequently underwent a one-sample t-test analysis to evaluate their correlation with each experimental condition. Some components (2, 4, 5, 12, 13, and 15 to 20) were identified as noise. The remaining components (1, 3, 6, 7, 8, 9, 10, 11, and 14) and their relation with the experimental conditions are described in **Table 28** (Appendix E) and Appendix F.

Two components significantly differed in their correlation with the experimental groups: Independent Component 1, identified as a set of regions involved in the visual scene analysis, was more active in the social group compared to the non-social group during subject extraction. This result was comparable across sentence context levels (same and different; **Table 3a**).

Component 20, identified as a set of areas typically involved in social cognition tasks, was surprisingly more active in the non-social group compared to the social group. However, this difference was observed only in the checkerboard observation phase, and only when words were repeated in a consistent context (**Table 3b**).

Table 3 – Correlation: independent component - experimental conditions

Results of the two-sample t-test testing for correlations between each independent component and the experimental conditions across the two groups (S+, S-; $p < 0.05$, FWE).

Table 3a – Differences between social and non-social groups for Component 1

	Independent Component 1: Analysis of the Visual Scene			
Experimental condition Non-Social ≠ Social	**P value**	**T value**	**S+ mean**	**S- mean**
cSC	n.s.	--		
cDC	n.s.	--		
sSC	0.023646491	-2.3550489	6.77 ± 0.86	6.11 ± 0.93
sDC	0.037485939	-2.1539144	7.54 ± 0.95	6.88 ± 1.03
oSC	n.s.	--		
oDC	n.s.	--		

Table 3b – Differences between social and non-social groups for Component 20

	Independent Component 20: Social Cognition			
Experimental condition Non-Social ≠ Social	**P value**	**T value**	**S+ mean**	**S- mean**
cSC	0.021893491	2.3877482	0.71 ± 1.08	1.47 ± 0.94
cDC	n.s.	--		
sSC	n.s.	--		
sDC	n.s.	--		
oSC	n.s.	--		
oDC	n.s.	--		

4.5.6 fMRI results, Part IV: DCM

We estimated the intrinsic connections for the condition sSC and dSC in a model including three volumes of interest in the right hemisphere: First, a region corresponding to the primary/secondary visual area in the calcarine gyrus, at the boundary between Brodmann areas 17 and 18 (V1); second, the right angular gyrus in correspondence with BA39 (rAG); third, the right middle frontal gyrus in correspondence with BA8 (rMFG). The three regions were fully connected except the middle frontal gyrus and V1. The entry point of the input to the system was supposed to be the angular gyrus. A schematic representation of the model is provided in **Figure 26**.

All estimated intrinsic connections in the model were significantly different from zero (all ps < .006, FDR-corrected), except in the SC-DCM model, the forward connection from V1 to the right angular gyrus. Results are summarized in **Figure 43** and **Table 4**.

At the second level, corresponding to a random-effects analysis at the group level, we performed separate ANOVAs for each of the connections, including the factors **run** (1, 2, 3), **context** (sSC or dSC) and **social interaction** (S+, S-).

The intrinsic connection ***V1 → rAG*** was not influenced by any of the experimental factors (all ps > .102). The opposite connection ***rAG → V1*** was stronger in the sSC (M = 1.325, SEM = .089) compared to the dSC (M = .992, SEM = .079) condition [F(1,39) = 11.104, p(unc.) = .002, p(FDR) = .008, η_p^2 = .222]. Further, the strength of the connection decreased during the experiment [linear trend, F(2,78) = 4.376, p = .016, η_p^2 = .101; Run1: M = 1.315, SEM = .074; Run2: M = 1.063, SEM = .097; Run3: M = 1.096, SEM = .087], although this result was only marginally significant when correcting for multiple comparisons (p = .064). Lastly, a main effect of group was evident, with the social group (M = 1.361, SEM = .095) having a stronger connection than the non-social group (M = 1.361, SEM = .097) F(1,39) = 8.884, p(unc.) = .005, p(FDR) = .020, η_p^2 = .186]. There were no significant interactions (all ps > .349). The intrinsic connection from the right angular gyrus to the middle frontal gyrus (***rAG → rMFG***) was modulated by sentence context [F(1,39) = 6.821, p(unc.) = .013, p(FDR) = .026, η_p^2 = .149]; more specifically, connections in the dSC condition were significantly stronger (M = .537, SEM = .057) than in the sSC condition (M = -.136[23], SEM = .034). Further, the strength of the connection significantly decreased over time [linear trend, F(2,78) = 3.303, p(unc.) = .042, η_p^2 = .078; run1: M = .276, SEM = .033; run2: M = .193, SEM = .036; run3: M = .134, SEM = .051]. However, this result did not survive correction for multiple comparisons [p(FDR) = .084]. There were no other effects or significant interactions (all ps > .126). The opposite connection, ***rMFG → rAG,*** was not influenced by any of the experimental variables (all ps > .061).

Table 4 - Intrinsic connections within the chosen network.

a) Mean coupling parameters for each session in the condition dSC: Mean, standard deviation (SD) and two-tailed statistical significance (p-value), FDR-corrected, are reported. b) Mean coupling parameters for each session in the condition sSC: Mean, standard deviation (SD) and two-tailed statistical significance (p-value), FDR-corrected, are reported.

4a) dSC-DCM

Intrinsic Connections	Mean	SD	P-Value (FDR)
V1 → rAG	.0446	.099	**.006**
rAG → V1	.9969	.538	**4.40E-014**
rAG → rMFG	.538	.366	**2.09E-011**
rMFG → rAG	-.0994	.190	**.002**

4b) sSC-DCM

Intrinsic Connections	Mean	SD	P-Value (FDR)
V1 → rAG	-0015	.150	0.949169
rAG → V1	1.329	.602	**1.49E-016**
rAG → rMFG	.705	.500	**7.03E-011**
rMFG → rAG	-.137	.215	**2.95E-004**

---

[23] The negative mean may imply that in the case of sSC, rAG actually depresses activity in the rMFG.

Table 5 – Effects of the experimental manipulations on the intrinsic connections

P-values for significant main effects are indicated according to the FDR correction for multiple comparisons; uncorrected p-values are indicated in parenthesis.

	Context		Run		Social	
	Direction	**P-value**	**Direction**	**P-value**	**Direction**	**P-value**
V1 → rAG	--	.136 (.102)	--	.277 (.369)	--	.441 (.441)
rAG → V1	sSC > dSC	**.008 (.002)**	1 > 2 > 3	**.064 (.016)**	S+ > S-	**.02 (.005)**
rAG → rMFG		**.026 (.013)**	1 > 2 > 3	**.084 (.042)**	–	.265 (.199)
rMFG → rAG	--	.333 (.333)	--	.739 (.739)	--	.186 (.093)

Figure 43 – Strength of intrinsic connections.

Strength of intrinsic connections is represented for sSC-DCM (left) and dSC-DCM (right). Solid lines represent intrinsic connections significantly different from zero; dotted lines represent connections not different from zero

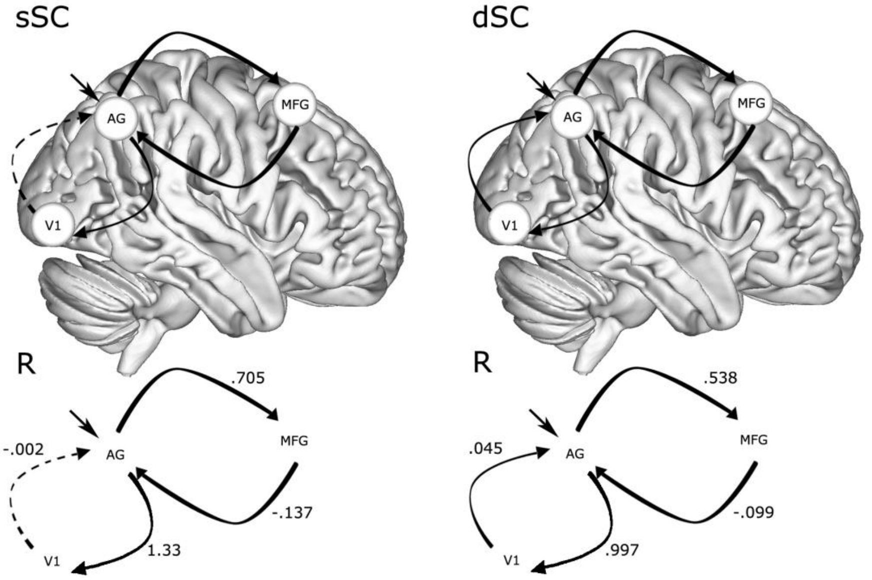


Figure 44 – Intrinsic connections: social interaction.

Intrinsic connections showing a main effect of social interaction (stronger connection in the S+ as compared to the S- group).

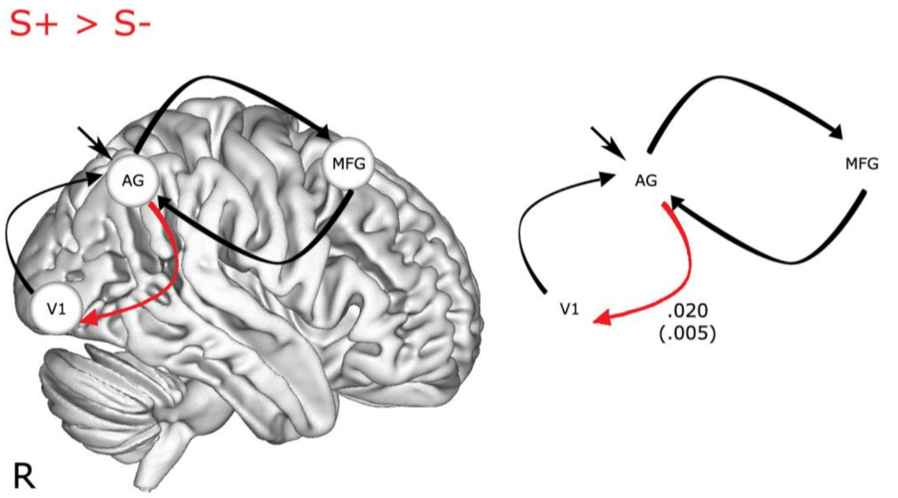


Figure 45 – Intrinsic connections: sentence context.

Intrinsic connections showing a main effect of sentence context (stronger connections for sSC compared to dSC)

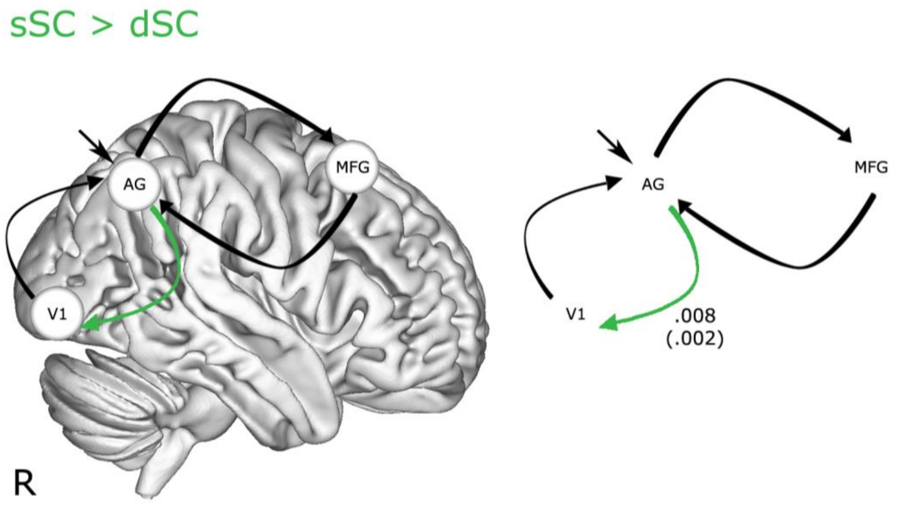
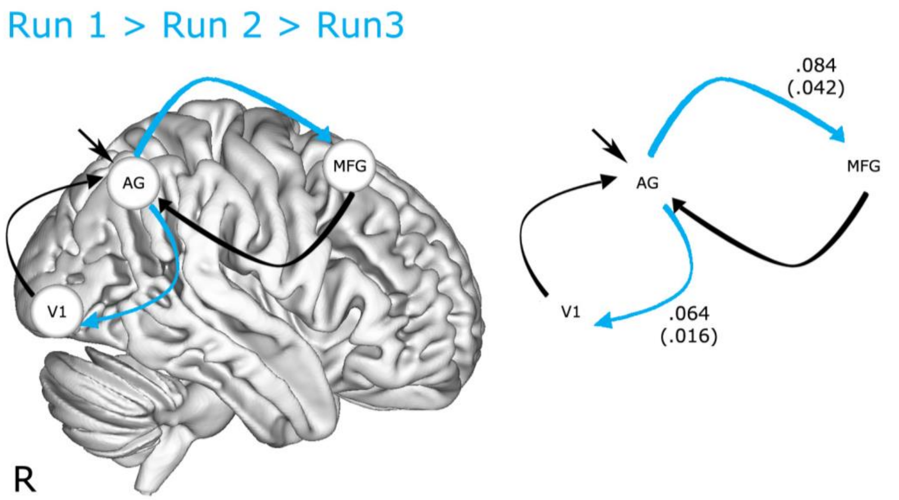


Figure 46 – Intrinsic connections: learning effects.

Intrinsic connections showing a modulation from beginning of the experiment to the end.



4.5.7 Discussion

The current study aimed to provide a comprehensive account of the neural substrates involved in contextual adult word learning during social interaction. Our experimental setting reproduces an ecological learning situation, in which participants believe that they are interacting with a knowledgeable partner; as such, this study goes beyond previous evidence for social language learning (for example, Jeong et al., 2010) by investigating social interaction via a "second person" approach, in which the learner is actively involved with a knowledgeable partner (Schilbach, 2014; Schilbach et al., 2013). Our results provide an extensive characterization of the influence exerted by social interaction and sentence context variability on verbal learning; more specifically, we suggest that social interaction significantly influences word learning by modulating the activity of task-relevant areas, such as the parietal lobe and visual areas in the occipital and infero-temporal lobe that are involved in visuo-spatial attention and word learning. Further, our results show that the intrinsic connectivity between these areas is enhanced by social interaction, even if participants only believe that they are interacting with a human partner, while they are in fact performing a computer game. Importantly, the extent to which social interaction influences word learning critically depends on task demands.

In the following paragraphs, we summarize and discuss the processes underlying different phases of the learning process and how social learning differs from non-social learning.

4.5.7.1 Checkerboards: Exploring the learning environment

In natural word learning situations, new words are frequently encountered and incidentally acquired from context (Jeong et al., 2010; Laufer & Hulstijn, 2001; Nagy et al., 1987; Rodríguez-Fornells et al., 2009; Swanborn & De

Glopper, 1999). When faced with an unknown word, an adult learner is similar to a child that has to scan the environment in search for possible referents. In the present task, this moment was represented by the appearance of a checkerboard display in which target referents were hidden. We did not expect to observe differences between social and non-social groups at this stage; indeed, this phase represented an initial step in the analysis of the visual environment, in which the knowledgeable partner was not actively supporting or helping the learner. Nevertheless, "socially" trained participants displayed activation of the left posterior hippocampus, a structure of the medial temporal lobe critically involved in learning and creating long-term memories (Bray, 2014; Breitenstein et al., 2005; Brown & Stern, 2014). This result may suggest a modulation exerted by social interaction upon task-related areas. However, these results emerged only at a lowered threshold, thus calling for further studies to provide replications at more conservative thresholds. This reduced statistical signficiance may be related to insufficient power due to the between-subjects design; in addition, there was no real difference in the stimulation that participants of the social group were exposed to, compared to participants in the non-social group. Indeed, both groups played exactly the same computer game. A more in-depth explanation of these two hypotheses is provided below.

To confirm that the appearance of the checkerboards represented an initial explorative moment, activations at this stage were elicited in areas involved in both low-level (Independent Component 1) and higher-level (Independent Component 2) visual analysis; further, checkerboards were positively correlated with the Attention Network (Independent Component 9) and negatively correlated with the Default Mode Network (DMN, Independent Component 14; Raichle et al., 2001), a set of areas recruited when cognitive effort is required. However, differences in the way checkerboards were processed emerged over the course of the experiment: At the beginning compared to the end of the experiment, the checkerboard analysis was supported by an extensive set of areas involved in visual exploration and cognitive control (anterior and middle portions of the cingulate gyrus, the inferior parietal cortex and the insula; for example, Conci & Müller, 2012), together with posterior peri-sylvian regions involved in language-related tasks and, in particular, with language comprehension (temporal and angular gyrus; Binder, Desai, Graves, & Conant, 2009). In contrast, at the end of the experiment compared to the beginning, greater activation was found in primary and secondary visual areas. Taken together, these results suggest that, at the beginning of the experiment, greater cognitive effort was required to understand and to decode the checkerboards, while at the end of the experiment, recognition thereof was largely delegated to low-level areas dealing with visual analysis. Alternatively, it is possible that participants learned to suppress the unnecessary information provided by the checkerboard and refrained from exploring the visual stimuli, at least when checkerboards were the same each time they occurred. Indeed, checkerboards depicting variable sentence contexts elicited greater activity bilaterally in the fusiform gyrus and the right cerebellum, than constant checkerboards. The fusiform gyrus is a high-level visual associative area, dealing with the integration of complex images and word reading (Thoma & Henson, 2011; Zhang et al., 2014); hence, this result suggests that a more complex visual analysis was employed to explore constantly changing environments. Crucially, activity in the fusiform gyrus during learning was significantly correlated with accuracy in the behavioral test. However, no brain areas were more activated by same context compared to different context checkerboards; yet, same context checkerboards were significantly correlated with the language network (Independent Component 7). This result is in line with our hypotheses that checkerboards depicting the same sentence context would be immediately recognized as previously learned items, thus triggering a process of anticipating the sentence context constituents. In this condition, the learner does not need a social partner to disambiguate possible referents; accordingly, we found activity of the network underlying social cognition (Independent Component 20) to be higher in the non-social compared to the social group during the presentation of same context checkerboards. We expected social interaction to be maximally important when task requirements were more difficult, as the learner may benefit from engagement with a knowledgeable partner (Verga et al., in prep.; Verga & Kotz, in prep.; Verga & Kotz, 2013); this situation may be represented by a constantly changing environment such as the one depicted by different context checkerboards. Conversely, social interaction may be redundant when the task is easy enough to be performed without external help, which was the case with the same context checkerboards. Hence, in this learning situation, participants in the social group may "suppress" the presence of the other person, resulting in reduced activation of the social cognition network.

Lastly, we observed a negative correlation of activity in the left anterior cingulate cortex with temporal coordination between the participant and the timing of the sentence context presentation by the computer; in other words, the more participants were temporally coordinated with the timing sequence of the computer (despite the "social" group thinking of a human partner), the less the ACC was activated. Since the ACC is recruited in decision making tasks, error monitoring, and cognitive control (Shenhav, Botvinick, & Cohen, 2013), the negative correlation may indicate that smoother temporal coordination with the timing sequence leads to decreased conflict; in other words, less effort is required to anticipate and compute the next step in one's own behavior. This interpretation is in line with our previous suggestions regarding the role of temporal coordination in word learning (Verga et al., in prep.; Verga & Kotz, in prep.; Verga & Kotz, 2013).

4.5.7.2 Subject and verb extraction: Creating the sentence context

After having explored the new environment represented by the checkerboards, the learner was presented with the sentence context. The type of processing involved in this phase was strongly influenced by the type of sentence context: Consistent (same) sentence contexts elicited widespread activations in a bilateral set of areas, involving the right cingulate cortex and putamen, bilaterally the supramarginal gyrus, superior temporal gyrus and cerebellum, and left inferior parietal lobule, amygdala and insula. These areas are often found together in tasks dealing with spatial attention and planning of action execution, including covert speech production (Nardo, Santangelo, & Macaluso, 2014; Thoma & Henson, 2011). This result is in line with conclusions drawn from the checkerboard observation stage; more specifically, since the sentence context could already be identified at the presentation of the checkerboards, during this next step the participants could focus on what to do next – locate and select the target. However, when the sentence context changed with each repetition of a novel word, activation increased in areas involved in the retrieval and integration of syntactic and semantic information (inferior frontal gyrus, bilaterally), together with areas involved in spatial analysis (calcarine gyrus, occipital gyrus) and areas involved in the identification and naming of visual stimuli (lingual and fusiform gyrus). This pattern of activation is consistent with the interpretation that participants worked towards the construction of the sentence meaning. Taken together, these results support our hypotheses that when a context is constantly repeated, the decoding of a sentence is not necessary, but if a context is always changing, the sentence context has to be decoded each time it occurs. However, how does social interaction impact this difference?

Even in this phase of learning, social interaction modulates the activity of task-relevant areas; the middle temporal gyrus and the left inferior temporal gyrus were found to be activated more in the social group than in the non-social group. These areas have been implicated in several tasks such as spatial analysis and object recognition, as part of the ventral visual stream (Thoma & Henson, 2011), and were found to be more active in variable sentence contexts. The right angular gyrus has been consistently reported in studies on complex social functions such as Theory of Mind (Carter & Huettel, 2013), but also in visuo-spatial attention tasks as a junction between the frontal regions (Han et al., 2004; Verghese et al., 2014) and the visual cortex (Constantinidis et al., 2013; Macaluso et al., 2000; Saalmann et al., 2007; Verghese et al., 2014). It has been suggested that the angular gyrus may re-direct attention towards new targets (Corbetta et al., 2000), ultimately facilitating the processing of stimuli at attended locations (Bressler et al., 2008; Thiel et al., 2004). In line with this interpretation, Independent Component 1 (visual analysis) was significantly more involved in the construction of the sentence context in the social group than in the non-social group. This result confirms one of the findings from the main analysis; that is, greater activation of the middle occipital and inferior temporal gyri in the social compared to the non-social group. Taken together, these results suggest that social interaction modulates the activity of task-related areas; more specifically, social interaction is used when the task is more challenging (variable sentence context) to increase the allocation of attention towards a target stimulus via the top-down modulation of visual cortices.

To confirm this hypothesis, we performed a DCM analysis to investigate intrinsic connections existing between areas involved in the visuo-spatial attentive network identified in the main analysis: The middle frontal gyrus, the angular gyrus, and primary/secondary visual areas. We hypothesized that social interaction during learning would increase activation within the TPJ and its connectivity with other task-relevant areas (Decety & Lamm, 2007; D. V. Smith et al., 2010). This hypothesis has a strong evolutionary and ontogenetic basis; indeed, a possible role for the attentional network during social interaction has been proposed for first language learning (Kuhl et al., 2003; Waxman & Gelman, 2009) and second language learning (Verga & Kotz, in prep.). Further, it has been shown that persons diagnosed with Autism Spectrum Disorders may display altered connections in the attentional network, with attention failing to modulate connectivity between extrastriate areas and V1 during a visual task

(Bird et al., 2006). Indeed, the results of our DCM analysis suggest that the top-down connection from the right angular gyrus to visual areas in the occipital lobe is reinforced by social interaction.

4.5.7.3 Object Encoding: Mapping words and referents

As a last step, after having investigated the environment and understood the sentence context, learners were faced with the task of binding the new word with its referent. Similarly to the checkerboard appearance, we did not expect, at this stage, that social interaction would influence this process; indeed, the task of reading and encoding a new word is rather individual. Accordingly, we did not observe differences between the two groups.

During this phase of learning, there was virtually no difference between words belonging to different conditions; nevertheless, new words corresponding to objects that followed a variable sentence context, elicited activations in a widespread bilateral network encompassing the inferior frontal gyri (with a bigger extension in the left hemisphere), the pre-central gyrus, the middle occipital gyrus, together with the left inferior temporal and occipital gyri, the right middle frontal gyrus, and the thalamus. These areas have previously been reported to be involved in word reading and learning (Mestres-Missé et al., 2008; Ye, Mestres-Missé, Rodriguez-Fornells, & Münte, 2011); however, in these studies, participants were extensively trained outside the scanner. In this study, the training phase was extremely quick, to ensure that the entire learning process could be captured during fMRI scanning. The fact that these areas were more activated in the variable compared to the same context condition, further suggests that the modality of encoding in the same context may be more global than in the variable context condition. That is, words presented in the same sentence context may be encoded together with the remaining sentence context in a global fashion, as signs of activity already "disappear" at the presentation of the subject (or agent) of the sentence context.

Conversely, words encoded in a consistent context did not elicit more activation in any area, than words following more variable contexts. Behavioral test scores were positively correlated with activity in the left supramarginal gyrus, a key area for word learning and, in general, language comprehension (Moore & Price, 1999; Price, 1998).

4.5.7.4 Open issues

The present study leaves open some issues concerning the lack of behavioral differences between the two groups of participants (social and non-social) during the learning phase, and the low level of statistical significance of the results attained in comparisons of the groups.

The pattern of behavioral results across the three learning phases (training, learning, testing) is particularly interesting: At first glance, the results during the learning phase differ when compared to our previous experiments, where participants were found to be significantly faster, less variable, and more temporally coordinated with a social partner compared to a computer. Instead, in the current study, no behavioral differences were found during learning. However, significant differences between the social and the non-social group emerged during the training phase and the testing phase, both taking place outside the scanning session. It must be pointed out that the current study was set up slightly differently to the behavioral studies, as we had to adapt the social situation to an fMRI setting. In the previous studies, both the experimenter and the participant were slower to perform the task at the beginning of the experiment, but became faster towards the end; this learning effect was simulated in the computer task by having similarly decreasing presentation times for the subject of the sentence context. However, no such ploy was used in the fMRI set-up to avoid systematic confounds in the bold signal estimation; in other words, it would not have been possible to differentiate between the effects of the systematic increase in the velocity of presentation from a true learning effect. For the same reason, participants had limited time to provide an answer, while they did not experience time constraints in the previous studies. Although the time provided was clearly sufficient to give an answer, the introduction of a time constraint possibly shifted the participants' focus of attention to this aspect of the task, leading to a less naturalistic type of interaction. This hypothesis is in agreement with the finding that when participants were free to decide their pace during both the training and the testing phase, a significant difference in behavioral performance emerged between the social and non-social groups. This latter point introduces the last difference between the current and our previous studies: This time, while participants all believed the cover story, they did not see the experimenter who was in another room. In previous studies, however, the experimenter was sitting side-by-side with each participant. Thus, it may be the case that the direct presence of the "teacher" is required to elicit observable behavioral changes in the learner. This would explain why behavioral differences emerged in the training and testing phases, that were both conducted outside the scanner room either together with (training phase) or in close proximity to (testing phase) the experimenter. Notably, the results of the training phase confirm our previous findings, as participants in the social group were faster and less variable than those of the non-social group despite the very short length of this phase.

More puzzling is the finding of slower response times for the social compared to the non-social group during the testing phase, especially because no differences in accuracy scores complemented these findings. Some insight into this result may come from the functional analyses. At a liberal threshold ($p < .001$, uncorrected), the social group displayed a higher correlation of activity in the right superior frontal gyrus with test scores, compared to the non-social group, during the encoding of a new word. The right SFG is an area involved in superior functions such as spatial cognition (Boisgueheneuc et al., 2006), and it is possible that this activation reflects the use of a different strategy by the social group to encode new words, in line with what has been proposed by other authors (Jeong et al., 2010); indeed, in the current study, we found that participants learning socially employed the right SMG (an area critically involved in word learning) more than participants learning non-socially. Importantly, the use of a different strategy does not necessarily imply differences in accuracy, as long as the task is easy enough for the two strategies to be equally effective; this would explain the lack of differences in accuracy scores during the testing phase. However, different encoding strategies most likely require different *decoding* strategies, such as a different route to access the semantically stored information. For example, if information encoded during social interaction required visuo-spatial attention, it might be possible that for decoding, one could use visuo-spatial attentive mnemonics to recall the stored information. In contrast, participants in the non-social condition, who had a more “direct” encoding, would also have direct access to the stored material; this would be reflected in differences in response times for words originally encoded with a different strategy, which we observed in the testing phase results.

Along the same lines is the evidence that, albeit at a lenient threshold of $p < .001$, the social group displayed higher correlations between activity in the left middle temporal gyrus at the checkerboard presentation, and lag-0 cross-correlations. The left middle temporal gyrus is involved in several tasks, such as extracting word meanings when reading, but also evaluating space and distances. In our previous studies, we hypothesized that temporal coordination (expressed by temporal lag-0 cross-correlations) could represent a strategy used in social interactive settings to create ‘common ground’, allowing the caregiver (or teacher) to direct the attention of the learner within the environment. Although we did not observe differences in coordination measures between social and non-social groups in the current study, the greater correlation observed in the social group suggests that temporal coordination during social interaction may indeed represent a way to create a ‘common ground’.

Altogether, these results suggest that the direct presence of another person may influence the strategy employed to encode new words; however, the extent to which the physical proximity with the social partner influence this process (see Kuhl et al., 2003 for a similar effect in first language learning) is open to further investigations.

Lastly, we wish to put forward a shortcoming of the current study. Brain activations due to social interaction appear only at a more liberal threshold of $p < .001$, uncorrected for multiple comparisons, or corrected using an extent cluster threshold. Two explanations (that are not mutually exclusive) are proposed: First, this result may be caused by insufficient power due to the between-subjects design. Second, it should be noted that the difference between the social and non-social group in the learning task was only due to a psychological manipulation; indeed, there was no difference in the stimulation that participants were exposed to in either group, as both groups played exactly the same computer game. Nevertheless, activations distinguishing the social from the non-social group are consistent with the previous literature, with the task characteristics, and with our initial hypotheses. However, further studies are required to corroborate these results.

4.5.7.5 Conclusion

In the current study, we evaluated the neural correlates of contextual word learning during social interaction. Our results suggest that social interaction influences the processes underlying word learning by modulating the activity of task-related areas. More specifically, we propose that the presence of a social partner modulates the activity of the network involved in visuo-spatial attention. It is possible that the input originates in the right angular gyrus given its role in multi-modal integration, which is required for a complex, multimodal “stimulus” such as a social partner. In addition, we suggest that this modulation is dependent upon the task specifics; if a task is easy enough, social interaction does not provide an advantage, and it is therefore ignored.

Taken together, these results provide novel insight into the mechanisms behind social word learning in adults, suggesting that the presence of a social partner may help the learner by directing her/his attention towards the correct referent in a given sentence context. Further studies are required to investigate how this effect is modulated by variables such as task complexity and the physical proximity (visual contact) between learner and social partner.

5

Discussion

"No man is an island"

John Donne

Learning a language during adulthood can be difficult. The classical view of language acquisition proposes that compared to children, who acquire their first language quickly and effortlessly, adults need to be taught explicitly, and will still never be able to attain native-like proficiency. As seen in the first chapter, however, years of research on bilingualism proved this point of view to be largely inaccurate, as even adult learners may master a language learned later in life.

The aim of this dissertation is to provide further insight into the mechanisms and processes, which may modulate word learning in adult speakers. More specifically, it is claimed that contextual variables such as social interaction may be critical to shaping the attained proficiency. As the extent to which social interaction influences L2 vocabulary learning has not been systematically investigated, this thesis draws on different fields (for example, social interaction, joint action, and social psychology) to provide a comprehensive account of this multi-faceted phenomenon.

Chapter 1 of this dissertation presented a theoretical background of the relevant literature concerning social interaction and word learning; more specifically, the first part of the chapter described evidence from the field of social cognition, which strongly suggests an influence of social interaction on human behavior. This claim forms the basis from which we can hypothesize that language learning may also be influenced by interaction, a hypothesis that has been extensively studied in children, but not in adult learners. Accordingly, in the second part of the chapter, a comparative approach confronting first and second language acquisition has been chosen, to highlight similarities and differences between these two processes. Evidence that largely shared processes underlie learning in both children and adults suggests a common learning mechanism, and justifies the extension of the known mechanisms of L1 learning (most notably social interaction) to L2 acquisition.

Chapter 2 presented the research questions explored in this dissertation. The main question this thesis aimed to answer concerns the role that social interaction plays in second language learning (***Question 1***). This modulation in learning may occur through temporal coordination spontaneously emerging in socially interactive situations (***Question 2***). Indeed, it has been suggested that interacting people may become "coupled" in their behavior, which, in turn, facilitates temporal coordination. However, it is unclear whether this coordination emerges only with a human partner or if it is a phenomenon occurring under other circumstances; for example, music, being a highly temporally structured stimulus, may also boost temporal coordination (***Question 3***). Either way, the emergence of temporal coordination is hypothesized to potentiate the attentional effect of a social partner (***Question 4***), possibly together with local properties of the linguistic input such as sentence context characteristics (for example, variability; ***Question 5***). The extent to which these dynamics are reflected in brain activity during word learning was explored in the final study (***Question 6***).

In chapter 3, a new method was described that allowed adult word learning to be tested in social contexts. This method represents a possible solution to complications caused by interactive settings in the study of high-level cognitive functions such as language learning. A validation of the setting was conducted in the two studies described in chapter 4. The other three studies presented in the same chapter employed the novel method proposed here to investigate word learning in healthy adults under different circumstances (with or without social interaction and music).

The evidence collected in chapter 4 allows us to conclude that social interaction is a modulating factor in word learning in adult speakers (***Question 1)***; however, the extent to which social interaction impacts vocabulary acquisition depends upon the characteristics of both the local linguistic context (for example, sentence context variability; ***Question 5***) and the features of the global context in which the task is performed (for example, visual contact). In particular, the presence of a social partner elicits spontaneous temporal coordination (***Question 4***), which may represent a mechanism through which social interaction modulates word learning (***Question 2*)**. Within this dynamic, a human partner represents a particular type of stimulus that participants coordinate with. Stimuli of a similar complexity and temporal structure to social interaction, such as music, do not elicit comparable coordinative behavior in the learning task here presented (***Question 3)***. At the neural level, social interaction significantly modulates the activity of areas involved in visuo-spatial attention and their connections, and influences the activity of areas involved in word learning (***Question 6)***.

The goal of this fifth chapter is to summarize the results of the experiments described in chapter 4, and to interpret them in light of the role played by social interaction in second language learning. Furthermore, limitations of the approach presented here are described, and possible clinical applications are provided in an overview. Lastly, based on evidence collected here as well as the literature reviewed in the first chapter, an integrative model of word learning in adults is outlined. This model suggests possible alternative routes in the word learning process, through which social interaction and linguistic context features may influence vocabulary acquisition during adulthood.

5.1 Summary of results

Is social interaction a shaping force in adult word learning? The studies presented in chapter 4 evaluated this hypothesis by employing a novel method, which was validated in an initial series of two studies. After this validation phase, the method was used in a series of three experimental studies. Taken together, the experiments presented here provide the first evidence of the role played by social interaction in word learning in healthy adults.

Study 1. 20 native German speakers were presented with 564 written transitive sentences (subject, verb and object). The participants' task was to indicate whether each sentence represented a plausible German expression, and whether the object of the sentence represented a good match for the sentence context. The vast majority of the sentences were evaluated as plausible; similarly, most objects were judged to fit well in their respective sentence contexts. Sentences and objects, which were judged as non-plausible or not matching, were either excluded from the sample or, when possible, modified to increase their plausibility.

Study 2. A set of three pictures depicting a subject, a transitive verb and an object was assembled for each of the 376 sentences resulting from Study 1. This new material was evaluated by a different sample of 20 German speakers who were required to judge the plausibility of each sentence and the goodness of the match between the objects and their respective sentence contexts. Lastly, participants were required to type in the sentence conveyed by the pictures. On average, both sentences and objects were judged as plausible and well-fitting. Furthermore, the sentences typed by participants corresponded to the intended meaning in the vast majority of tcases. The few sentences rated as non-plausible were modified and later re-checked informally by native German speakers. Pictures with a low rate of agreement between the intended and understood meaning were excluded from the database. The sentences evaluated in these first two studies were used for the learning-game employed in the three main experiments.

Study 3. The word-learning game was employed to investigate whether social interaction enhanced adult word learning, and to what extent it interacted with the linguistic properties of the sentences (that is, variability in the sentence context). 68 native German speakers were exposed to either a non-social or a social version of the word learning game. In both cases, the participants' task was to find an object that correctly completed the sentence context by exploring checkerboards containing several black-and-white drawings. If participants gave a correct response, a new word for the object was presented. In the social condition, the experimenter's task was to select the subject and verb of the sentence (that is, the sentence context); in the non-social condition a computer provided the same information. To evaluate the interplay between social interaction and linguistic context, each target object could be repeated either in a varied or in a consistent sentence context. Participants learning socially were significantly faster, less variable and more coordinated with the experimenter than participants learning with the computer. Furthermore, they displayed improved temporal coordination when learning words repeated in a varied sentence context, which were also recognized significantly more accurately during the testing phase. No such differences were observed for participants trained non-socially. It was concluded that social interaction modulates word learning in concert with contextual cues: While the social partner facilitates the referent identification, a varied context provides new information at each occurrence of the new word to enrich and strengthen the forming concept.

Study 4. Study 3 showed that participants learning socially tended to coordinate temporally with a partner, possibly because the presence of another person functions as an oscillator which, in virtue of its complexity and temporal features, is able to catalyze the learner's attention. The current study investigates the specificity of this effect by exploiting another known stimulus with oscillator properties – music. 80 native German speakers performed the learning game under one of two conditions – social interaction and music – and their possible combinations, in a 2 x 2 factorial design. Results replicated the findings from Study 3, showing that social participants displayed better temporal coordination– with a partner than with a computer, especially in the varied context condition. Temporal coordination to music was weaker, and significantly impaired by the concomitant presence of a social partner. In the testing phase, participants who learned non-socially recognized words learned in a varied context more accurately than words learned in a consistent context. These results suggest that music and social interaction influence participants' learning behavior, but may hinge on different mechanisms. This conclusion has important implications for situations such as music therapy, in which music and social interaction are present at the same time.

Study 5. The fifth and last study aimed to identify the neural mechanisms supporting word learning during social interaction in healthy adults, by focusing not only on areas involved in this process, but also on their reciprocal connections (effective connectivity). Results of this study suggest that social interaction influences the processes underlying word learning by influencing word-learning areas (middle and inferior temporal gyrus, inferior frontal gyrus, TPJ, and subcortical structures). More specifically, the presence of a social partner in the word-learning game modulated activity in the network involved in visuo-spatial attention (including the middle frontal gyrus, angular gyrus and visual cortices); most likely, the input originates in the right angular gyrus, possibly due to its integrative role that would be required for a complex, multimodal “stimulus” such as a social partner. This modulation is dependent upon the task specifics, such that if the task is easy enough (for example, with a repeated sentence context) social interaction does not provide an advantage, and it is therefore ignored. These results provide evidence that an attentional mechanism might explain the impact of social interaction on word learning in adults, similar to what has been suggested in first language acquisition.

In the following paragraphs, the evidence summarized in this paragraph is explored and interpreted in further detail.

5.2 Is social interaction involved in adult word learning?

The aim of this dissertation was to answer the question as to whether social interaction may be involved in adult word learning. This question is, surprisingly, still very open, as it only recently started to attract the interest of cognitive neuroscience. As described in paragraph 1.1, 3.1 and 4.5, this lack of interest has classically been motivated by two lines of reasoning: The first is theoretical in nature, and states that adults’ cognitive functions are largely independent and self-sufficient; for this reason, the presence of other persons is not supposed to influence processes occurring in one’s own mind (Schilbach, 2014; Schilbach et al., 2013; Stephens et al., 2010) . The second reason is mainly methodological: Socially interactive settings pose several issues in terms of the planning and implementation of experimental designs (chapter 3.1). The work presented here overcomes both of these caveats, by presenting a novel method to investigate adult word learning during social interaction in a well-controlled experimental setting. By using this method, the studies summarized in chapter 4 provide a significant advance in our knowledge of the processes underlying adult word learning. In the following paragraphs, the main results of the experiments are discussed in detail and grounded in the theoretical background proposed in chapter 1.

5.2.1 The other's influence: Temporal coordination

The experiments presented here show that during social interactions participants engage in a temporally coordinated behavior with their partner. Furthermore, this behavior emerges spontaneously during the task, as participants were not explicitly required to coordinate with the experimenter. The literature on joint action consistently reports that partners tend to fine-tune the timing of their actions, not only in 'simple' motoric tasks (finger tapping, Pecenka, Engel, & Keller, 2013; jumping, Vesper, van der Wel, Knoblich, & Sebanz, 2012; performing martial arts, Schmidt, Fitzpatrick, Caron, & Mergeche, 2011), but also during higher-level cognitive performance such as conversation (Shockley, Richardson, & Dale, 2009). In the latter case, temporal coordination improves communication efficiency by minimizing overlaps and long silences (Stivers et al., 2009; M. Wilson & Wilson, 2005).

However, why do people tend to coordinate with each other? It has been suggested that this tendency depends upon endogenous oscillators in the brains of a speaker and a listener becoming coupled (Dumas et al., 2010; M. Wilson & Wilson, 2005). Wilson and Wilson (2005) describe these endogenous oscillators as "timing devices", constituted by "populations of neurons that collectively show periodicity in their activity and serve timing-related functions". Evidence in support of this proposition comes from several studies that use novel, interactive set-ups (for example, hyper scanning techniques, paragraph 1.1.2), showing that the brain activity of people involved in social interactions indeed becomes temporally coupled (Cui et al., 2012; Dumas et al., 2011, 2010; Hasson, Nir, Levy, Fuhrmann, & Malach, 2004; Montague et al., 2002). At the behavioral level, this neural coupling is mirrored by the temporal coordination phenomena described in the joint action literature (Pecenka & Keller, 2011;

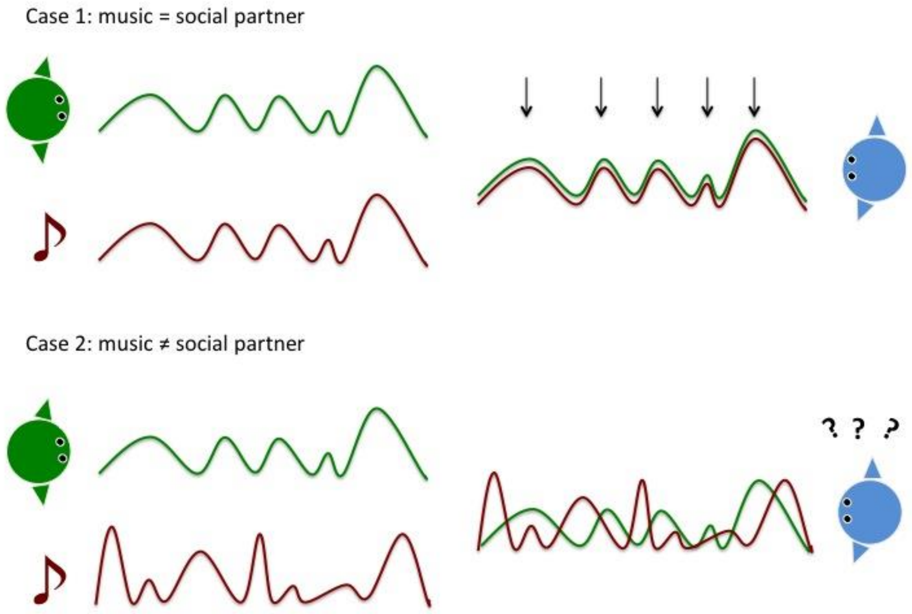
Pereira et al., 2008; Yun et al., 2012) and supported by results from studies 3 and 4 presented in chapter 4 of this thesis. This “coupling” between a participant and her/his partner has important consequences, as it establishes a “common ground” in which the exchange of information is facilitated (Csibra & Gergely, 2011; Tomasello & Carpenter, 2007). In the case of word learning, temporal coordination may represent a mechanism used to direct the learner’s attention towards a target referent for a new word at the correct point in time (Gogate et al., 2000; Rader & Zukow-Goldring, 2012; Rolf et al., 2009). In this way, the number of possible referents is strongly diminished and, consequently, word learning is facilitated (Csibra & Gergely, 2009; Hirotani et al., 2009). Accordingly, participants who learned socially in Study 3 recognized more words learned in the condition in which temporal coordination was better (that is, with varied context).

However, if the role of a social partner were simply that of an oscillator that facilitates attention orientation in time, other “oscillators” should impact word learning in a similar fashion. Study 4 tested this hypothesis by employing music as an oscillator that shares many of the same social interaction features; that is, it is a complex, temporally structured stimulus, often used to convey meanings. In particular, music’s temporal structure leads the listener to spontaneously coordinate with the sounds (Chen, Penhune, & Zatorre, 2008; see also Loehr, Large, & Palmer, 2011; Repp & Su, 2013) as a result of the extraction of temporal regularities (Jungers et al., 2002; Large & Palmer, 2002; M. Wilson & Wilson, 2005). This form of auditory-motor synchronization to music improves attentional processing, by facilitating the temporal encoding of stimuli (Schmidt-Kassow et al., 2013) such as new words (Francois & Schön, 2010; Jones & Boltz, 1989; Schön et al., 2008; Thaut, Peterson, & McIntosh, 2005; Ferreri & Verga, in prep.).

Hence, the rationale behind comparing music and social interaction is the following: If music and social interaction provide congruent temporal information (that is, the temporal structure helps to predict a subsequent event to occur in both cases at the same point in time), it should be maximally easy for the learner to plan her/his action when both sources consistently predict an event to occur at the same specific point in time (**Figure 47**, case 1). The observable outcome of this prediction would be similar temporal coordination occurring to either music or a social partner, and maximal temporal coordination to both. As a consequence, based on an attentional orienting account there should be no difference in word learning between the two conditions. The results of Study 4, however, demonstrate that music and social interaction elicit different behavioral outcomes in a learning task: Temporal coordination to music, compared to a human partner, is weaker and possibly more difficult to achieve. This result contradicts the hypothesis that music and social interaction facilitate word learning simply by improving attentional processing as a consequence of an improved temporal encoding of stimuli. Further support for this claim comes from evidence (from Study 4) that the concomitant presence of music and a social partner maximally hinders coordination. This result suggests that music and social interaction may provide different forms of information, making it difficult for the learner to integrate them and, hence, decide which external "time-keeper" to coordinate with (**Figure 47**, case 2). A similar conclusion has been reached by Demos and colleagues (Demos et al., 2012), who investigated spontaneous coordination to music and to a social partner: Pairs of participants were seated side by side in rocking chairs, either facing a partner, listening to music, or both. Results showed that coupling (temporal coordination) with music was weaker than with a partner, and that music competed with the partner's influence, thus reducing coordination. While this evidence is compatible with the idea that spontaneous coordination may derive from perceptuo-motor processes (described by the dynamics of coupled oscillators), it also supports the claim that music and a social partner provide different forms of information.

Figure 47 - Music and social interaction as oscillators.

Music and social interaction are hypothesized to represent oscillators, to which people may coordinate. But how similar are they? Case 1: Similar oscillators maximally favor coordination, by creating an overlapping congruous signal. Case 2: Music and social partner represent different types of oscillators, providing different information.



The fMRI study presented in chapter 4 provides, in this regard, very interesting and challenging evidence. On the one hand, the results of the training phase nicely replicate data from the previous experiments (that is, faster reaction times, reduced variability of reaction times), and the fact that these effects emerge despite the brevity of this phase suggests a strong tendency for humans to modify their behavior when interacting with others. On the other hand, this phenomenon did not emerge during the fMRI learning session. Both personal communications with participants and results of the paper and pencil questionnaires (Appendix C) testify that participants in the social group believed the cover story *("you will be playing with the experimenter, exactly as you did in the training")*; furthermore, most of the participants reported perceiving the "experimenter" either slowing down or speeding up in response to their behavior. These two pieces of information rule out the possibility that the lack of temporal coordination may be due to participants not believing the cover story. Two alternative explanations may be that a) visual contact with the partner plays a significant role in establishing coordination, and/or b) when not explicitly asked for, temporal coordination may be hindered by time-pressured interactions. These two possibilities are further explored and proposed as future directions of investigation in paragraph 5.3.1.

Despite a non-significant difference in the behavioral data, neural activity correlated with temporal coordination was significantly different between participants in the social compared to the non-social group. More specifically, the correlations between activity in the left middle temporal gyrus and left calcarine gyrus with temporal coordination were significantly higher in participants who learned socially. These areas are particularly relevant for the task at hand; more specifically, the middle temporal gyrus is a critical region for word learning (for example, Abutalebi, Cappa, & Perani, 2001; Mestres-Missé, Càmara, Rodriguez-Fornells, Rotte, & Münte, 2008; see also paragraphs 1.1.4 and 1.2.3.2), while the calcarine gyrus is a sensory area strongly implicated in visual perception and attention (Rossi & Pourtois, 2014; Verghese et al., 2014). Furthermore, evidence of different brain activity between conditions in the absence of behavioral correlates is not uncommon (Francois & Schön, 2010; Jäncke & Sandmann, 2010), as imaging techniques are often more sensitive than behavioral measures to subtle phenomena. Evidence from the behavioral data in Study 5 which showed the same trends as in studies 3 and 4, albeit not reaching significance, supports this possibility: The introduction of a time-pressured response may have reduced temporal coordination, so that its effect needed a more sensitive measure to be detected.

Taken together, these results are important *per se* to demonstrate that adults' behavior is significantly influenced by the presence of others (see paragraph 5.4.1 for possible modulating factors). However, they also open up the possibility that spontaneous dynamics emerging during social interaction may support communicative functions in adult learners. While similar indications have emerged from studies on first language learning in both healthy (Louwerse et al., 2012; Pereira et al., 2008; Rader & Zukow-Goldring, 2012) and clinical populations (for example, children suffering from autism; Kasari, Sigman, Mundy, & Yirmiya, 1990; Mundy, Sigman, & Kasari, 1990), the studies presented in this dissertation provide the first evidence that this phenomenon may occur in second language learning. The following paragraphs will explain and discuss evidence from the studies presented in chapter 4 that shows that this may indeed be the case.

5.2.2 The social effect: An attentional spotlight

As previously discussed, the experiments summarized in chapter 4 consistently show that social interaction induces spontaneous temporal coordination. This effect resembles the behavior of entrained oscillators, and is possibly specific for social interaction compared to other forms of temporally structured stimuli, such as music. It was hypothesized that temporal coordination facilitates learning by orienting the learner's attention towards a new word's referent at the correct point in time (Gogate et al., 2000; Rader & Zukow-Goldring, 2012; Rolf et al., 2009), thus greatly reducing the number of possible referents (Csibra & Gergely, 2009; Hirotani et al., 2009). Accordingly, previous evidence in first language learning has shown that temporal coordination in caregiver-toddler dyads significantly correlates with vocabulary expansion (Pereira et al., 2008).

fMRI data collected in Study 5 suggest that this may indeed be the case, even for adult word learning. First, the social group of participants, compared to participants learning non-socially, displayed a significantly higher correlation between coordination with a partner and activity in areas involved in word learning and processing (caudate nucleus, lingual gyrus, middle temporal gyrus). Second, activations in areas involved in visuo-spatial attention (for example, middle frontal gyrus, angular gyrus and visual cortices) were greater in the social group of participants. This evidence is further corroborated by the analysis of effective connectivity between these regions, showing stronger connections in the social compared to the non-social group. Third, social learners gave faster responses than participants learning non-socially (Study 3, Study 4, and Study 5 training phase). It may be argued that this effect merely reflects a "speeding up" due to social facilitation (chapter 3.1); however, social facilitation usually occurs when people are being observed, but not when they interact with someone. Furthermore, a simple increase in response speed would predict an increased error rate, which was not found in any of the experiments. Instead, this result is compatible with the interpretation that a social partner may orient the participant's attention towards the correct target referent, thus speeding up her/his responses.

Taken together, the data presented here suggest that a social partner may facilitate adult word learning by orienting the learner's attention towards relevant elements in the environment, not dissimilarly from what was proposed by most L1 learning theories (chapter 1.2.2). As an example, the social pragmatic theory (chapter 1.2.2.3) identifies joint attention between a learner and a caregiver as crucial in establishing a common ground, in which the child is able to determine the adult referent for a new piece of language without even knowing the language (Tomasello, 1992, 2000). While children are certainly facilitated by social interaction, in which caregivers usually employ simplified ostensive behavior and speech (chapter 1.2), temporal coordination and other social cues (for example, eye gaze, body posture and movements, etc.) also contribute to make a social partner a particularly salient and multi-modal "stimulus" (Kuhl, 2007; Sage & Baldwin, 2010).

The fact that social interaction conveys many different types of cue is supported by the recurrent activation of the right angular gyrus in social cognition studies, including Study 5 presented here. The role played by this region in multi-modal integration would be particularly helpful for a complex "stimulus" such as a social partner. In this light, consistent activations in the right parietal lobe reported by the social cognition literature may, in fact, be due to a general-purpose function of this area (Carter & Huettel, 2013; Decety & Lamm, 2007); that is, the right parietal cortex may serve a general function in directing attention, which would be recruited by social interaction. According to this proposal, "changing from a nonsocial to a social context should increase activation within the TPJ and its functional connectivity with other task-relevant regions; […] should the TPJ indeed be critical for establishment of that social context, then […] when social information becomes irrelevant for behavior, the TPJ should be disengaged even if a social agent is still present" (Carter & Huettel, 2013). The results of Study 5 support this prediction; in particular, the differences observed for words embedded in varied compared to consistent contexts are in line with the idea that the right parietal cortex should be disengaged when social information becomes irrelevant. This latter point is explored in detail in the next paragraph.

5.2.3 Interplay with the local context: Sentence context variability

The studies introduced in this dissertation investigated adult word learning starting from the assumption that the sentence context, in which a new word is embedded, may facilitate a learner in solving the indeterminacy of the referent problem. As described in the first chapter, language learners are able to derive the meaning of a new word from just a few exposures. In this process, each word repetition adds information to refine the corresponding concept; however, the extent to which the linguistic context of repetition influences the formation of a memory trace is still a matter of debate (Besson & Kutas, 1993; Hills et al., 2010; Perry et al., 2010; Steyvers & Malmberg, 2003; Waxman & Gelman, 2009). As of yet, evidence supporting different theories of human memory (Anderson & Bower, 1972; Waxman & Gelman, 2009) either shows a preference for a more consistent context (Dempster, 1987; Hicks et al., 2005; Koffka, 2013; Steyvers & Malmberg, 2003; Young & Bellezza, 1982) or the importance of context variability as a successful mnemonic device (Hills et al., 2010; L. B. Smith, 2000). In the first case, it has been proposed that words consistently presented in the same sentence context may be encoded as a unitary episode; hence, their mnemonic representation should be highly consistent, yet difficult to extend to novel contexts. On the contrary, sentence context variability is suggested to improve generalization to novel items (Perry et al., 2010).

As noted in chapter 2, it is hypothesized here that the advantage of one or the other type of encoding may be significantly different in a social compared to a non-social learning situation: If a knowledgeable partner helps the learner to identify the correct word referent, a varied context may improve learning by adding information to refine the referent's concept. However, in a non-social situation, solving the indeterminacy of the referent at every occurrence of a new word may be particularly difficult. Hence, in this case, a consistent context may be more beneficial to learning. This hypothesis allows specific predictions to be made: First, that a knowledgeable partner is particularly useful in the case of a varied context, and less so in a consistent context; second, that this should be reflected by higher temporal coordination, increased activity in areas critically related to the task and, lastly, better results in the testing phase for words embedded in a varied compared to a consistent context.

The results for temporal coordination confirm this prediction; in both Study 3 and Study 4, temporal coordination in the social group was higher in conditions in which words were embedded in a varied context. These results suggest that, in the social condition, help from a partner is more useful, as the possible referent for a new word is not known a-priori. In this case, the partner may direct the learner's attentional spotlight towards the correct word referent. Conversely, when words are constantly repeated in the same context, the partner is not needed to solve the indeterminacy of the referent. Accordingly, significantly increased activations were found in areas involved in word learning and visuo-spatial attention in the social compared to the non-social group in the varied context condition. In particular, effective connectivity analysis revealed greater modulation of the backwards connections from the right angular gyrus to primary visual areas in the varied (but not in the same) context condition, supporting the idea that the right TPJ may play a supra-modal role and disengage when social interaction is no longer necessary for behavior, even if the social partner is still present (Carter & Huettel, 2013).

The results of the testing phase for the experiments presented in chapter 4 provide a somewhat complex picture in relation to this issue, possibly related to the type of retrieval mechanism employed in the different tests. In Study 3, a recognition and a recall task were performed to evaluate learning. The recognition test revealed that participants who learned non-socially recognized words encoded in a consistent context better, while participants learning socially recognized words encoded in a variable context better. However, the recall task revealed an advantage of different context words regardless of the type of social context. In Study 4 and Study 5, a novel testing phase was used in which participants were required to complete a novel sentence context with one of the words learned during the learning phase. This type of testing represented a particular type of recall task, which allowed us to test for a generalization effect. Using this task in Study 4, participants who learned non-socially remembered more words originally encoded in different sentence contexts, while no differences were observed for participants who learned socially. Lastly, in Study 5, no differences were observed between groups; however, participants who learned socially were significantly slower than participants who learned non-socially. Taken together, these results suggest two possible conclusions: First, in the current learning task a varied context helps to strengthen and enrich a memory trace for a new word by providing elements to refine word-concept mapping (Hills et al., 2010; Perry et al., 2010; L. B. Smith, 2000). Second, recognition and recall tasks hinge on different aspects of the memory process; recognition is easier and does not require a deep level of encoding, while recall requires the access and retrieval of a stored item (Craik & Lockhart, 1972; Moscovitch & Craik, 1976). The fact that social effects emerge during recognition, but not during recall, may indicate that social interaction provides a low level of processing depth. Another possible explanation is that the difference in the learning and the testing phase (social – non social) may have hindered retrieval. This hypothesis also explains evidence from the fMRI study, in which social participants were slower than non-social learners during testing. This evidence contradicts previous findings, as participants learning socially are usually *faster* than participants trained on a computer (paragraph 3.1). However, as previously suggested, the learner performed the testing task alone, while the learning phase was conducted with a social partner. A different type of encoding with respect to the type of retrieval predicts slower reaction times, as the route to decode the stored information would be different. Alternatively, it has been proposed that slower processing times may be an index of deeper levels of encoding (Craik & Lockhart, 1972; Craik & Tulving, 1975). As deeper levels of encoding are associated with stronger memory traces, evidence of slower processing in the social group could indicate that words learned socially were encoded more accurately than words learned non-socially.

These two possibilities (learning – testing congruency and depth of encoding), their implications for the current results, and possible avenues for future investigations, are explored in more detail in paragraph 5.4.2.

5.2.4 Summary

The studies presented in this dissertation provide a significant advancement in our understanding of the processes underlying word learning in healthy adults during social interaction. Three lines of evidence have been summarized in the previous paragraphs: First, the presence of another person significantly influences the learner's behavior; in unconstrained settings, in which participants are not time-pressured to answer, this influence is reflected in spontaneous temporal coordination emerging between partners of a dyad. Second, a knowledgeable partner directs the learner's attention towards the referent for a new word, and third, the role played by a knowledgeable partner in adult word learning critically depends upon the specifics of the task; if the context allows the learner to identify the referent by her/himself, the presence of another person does not constitute an advantage.

The set of studies presented here, however, leaves several questions open that will require further investigation. In particular, three main avenues for future research are identified: First, the differences between the behavioral studies (Study 3 and Study 4) and the fMRI experiment (Study 5) in terms of temporal coordination, suggest that direct visual contact with a partner may be relevant for social learning. Alternatively, it is possible that spontaneous temporal coordination may be influenced by task instructions (for example, time pressure). Second, an important part of learning is to retain and to be able to re-use information over time; however, the current studies do not allow social effects on long-term memory to be traced. Third, differences in the testing phase in different experiments suggest that a closer look at the levels of encoding and the consistency of retrieval should be considered.

While these possibilities for future investigations mostly stem from caveats observed in the current studies, the proposed method surely has some important potential for clinical applications, which will be explored in an overview.

5.3 Caveats, open questions and future directions

While the present dissertation provides answers to several questions concerning the role of social interaction in adult word learning, some caveats must be put forward and several queries still remain open. In both cases, more in-depth future investigations are required. In the following, caveats of the current research are outlined, together with possible solutions and proposals for future lines of research, which may help in clarifying aspects left open by the current studies.

5.3.1 Role of visual contact and instructions

A possible critique to the outcome of the current set of studies may be that temporal coordination during the learning phase clearly emerged in Study 3 and Study 4, but was not evident in Study 5, which was conducted in an fMRI setting. However, in the same study, indexes of temporal coordination (faster reaction times, reduced variability) emerged in the training phase. Why was this the case? A difference between the learning phase of Study 5 and the other studies was that, in the fMRI study, participants were not in the same room as the experimenter. This set-up precluded the possibility for the learner to actually see her/his partner, as only task stimuli were visible via the mirror inside the scanner. As a consequence, the possibility that temporal coordination may be modulated by visual contact with the social partner needs to be explored. In fact, typical joint action situations are characterized by direct visual contact between the partners (Bigelow, 2003); for example, in joint attention situations eye gaze represents a pivotal cue to infer and/or direct the other's attention towards a relevant new element in the environment (M. Carpenter et al., 1998; Saito et al., 2010; Tomasello, 1995). More generally, visual observation is deemed important in many socially interactive settings (Oullier et al., 2008), as observing others' actions may help in understanding what the other person is doing, in turn facilitating the planning of one's own course of action (Nummenmaa & Calder, 2009). This ability largely depends on the activation of the mirror circuit in the brain (Ciaramidaro et al., 2014; Hamilton, 2013; Oberman & Ramachandran, 2007). Despite the fact that all participants understood that they were interacting with the experimenter, in the present fMRI setting the interaction was not face-to-face. Hence, cues that may have facilitated coordination in the other settings (such as seeing the experimenter start a movement to select the right element on the checkerboard) were not available to the participant. A possible avenue for future studies is to explore the extent to which visual contact between people may influence their joint action performance. The answer to this question is of particular relevance considering, for example, the increasing availability of web-based learning opportunities, such as on-line courses offered by many universities. If face-to-face, direct visual contact is proven to be pivotal for learning, the way these learning opportunities are structured needs to be critically re-evaluated. A possible way to pursue this hypothesis would be to test participants either with the standard game-learning setting (for example, side-by-side in front of a computer), or with a setting in which a panel placed between the experimenter and participant precludes the possibility of actually seeing the partner. It is important to point out that further modification to the standard setting may allow progressively more detailed information to be obtained; for example, both of the proposed conditions ("standard" learning game and "visually blocked" learning game) may be tested with earplugs to specify the extent to which auditory feedback in addition to vision contributes to social coordination.

Another possible explanation for the difference in coordination across studies relates to the instructions given to participants. Indeed, participants were only given a specific time limit for their responses during the fMRI learning phase. It is possible that this time-pressure shifted participants' focus from interaction with the experimenter to their own performance. Instructions on how to perform a task are well known to influence behavior (Eiriksdottir & Catrambone, 2011; Schneider, Nott, & Dux, 2014). A classical example comes from evidence that it is possible to modify the behavior of a participant in a speed-accuracy task by asking her/him to focus either on giving as many correct responses as possible (accuracy) or on giving answers as fast as possible (speed; Heitz, 2014; Reed, 1973). The question thus arises as to whether the type of social coordination observed in the present studies may emerge only (or predominantly) in "natural" interactive settings, in which no time pressure is posed on the participants. A possible way to test this hypothesis would be to directly present participants with the same learning task but with different instructions: In one case they are required to answer as fast as possible, while in the second case they do not have a time limit to answer. This type of study would allow a direct comparison of the learning processes explored in Study 3, Study 4, and Study 5, and could have important implications for learning in everyday life. While most natural adult learning situations are instances of temporally unconstrained interactions, the effect of time-limits on coordinative performance may shed light on the effect of pressures on social learning. The school classroom provides an example of a context in which both the learning as well as the testing of acquired knowledge usually has specific time constraints. If social learning is influenced by time pressure, this type of teaching style may need to be revised.

5.3.2 Social interaction and memory

The studies presented in this dissertation all examined the effects of social interaction on word learning. While this is certainly the first step towards acquiring a new language, an important additional step is to maintain the newly acquired information over a longer period of time. In other words, while it is important for the learner to understand that “gavagai” stands for rabbit, it is equally important for this information to be consolidated. If the learner does not remember what “gavagai” means a week later, the learning process will have to start over again and this would be very inefficient.

While the studies put forward in this dissertation focused on learning, it is undeniable that future research should also explore the extent to which social interaction may impact memory of words. In the following paragraphs, two avenues are proposed to address this question. The first suggests the use of a testing phase congruent with the modality of the encoding phase in order to identify the possible generalizability of social effects and, hence, the relative strength of memory traces created in either of the two conditions. The second evaluates the impact of social interaction on long-term memory traces by proposing a delayed retrieval phase that allows testing for memory consolidation effects.

5.3.2.1 Levels of processing versus transfer appropriate processing

What does the strength of a memory trace depend upon? Studies on learning and memory have established that the possible factors involved in determining the strength of encoding are the level at which the processing occurred (Craik & Lockhart, 1972), and the characteristics of the encoding context (for example, the complexity of the context a new word is embedded in; Craik & Tulving, 1975). According to this view, greater “depth” implies a greater degree of semantic or cognitive analysis (Craik & Lockhart, 1972); items processed at a deeper level of analysis (for example, semantic) create a more persistent and elaborate memory trace compared to items processed at a more superficial level (for example, phonological). Within this framework, social interaction may create strong memory traces by allowing a deeper processing of words. The rationale behind this hypothesis (as proposed in the introduction of this thesis and supported by the present studies), is that a social partner helps the learner by directing her/his attention towards relevant elements in the environment (for example, Csibra & Gergely, 2009; Verga & Kotz, 2013); as a consequence, the allocation of attention to the correct referent (for example, the word meaning) determines a deeper level of processing, resulting in a more stable memory trace (Stein, 1978). However, in the current experiments, a difference between retrieval of words encoded socially or non-socially was not observed which does not support this interpretation.

An alternative explanation comes from the framework of the *transfer appropriate processing theory*. This approach states that the strength of a memory trace (that is, the ease of its retrieval) depends, more than on the level of processing, on the type of encoding and the type of retrieval (Stein, 1978; Tulving, 1979): If the modality of encoding is congruent with the modality of testing, retrieval should be facilitated. This approach suggests an encoding specificity, and emphasizes that the type of learning should be defined relative to how the acquired information will be used (Morris, Bransford, & Franks, 1977). In the studies described in chapter 4, the social groups faced an incongruity between the learning phase, in which participants were learning socially, and the retrieval phase, which they always performed alone. Thus, non-social groups were exposed to congruent types of encoding and testing (that is, alone with a computer). According to the transfer appropriate processing approach, participants in the non-social group were likely facilitated during the retrieval phase, while the social group was at a disadvantage. An explanation based on incongruence between the type of encoding and the type of testing has been put forward to explain results of other learning studies; for example, Peterson and Thaut (Peterson & Thaut, 2007) found no behavioral advantage for sung compared to spoken word lists in an explicit learning task (Rey's Auditory Verbal Learning test), in which words were sung during learning and spoken during the recall phase. However, the same authors found a behavioral advantage for sung stimuli when participants were instructed to sing back during the recall phase (Thaut, Peterson, McIntosh, & Hoemberg, 2014; Thaut et al., 2009; Ferreri & Verga, in prep.). The possibility that a transfer appropriate processing effect explains the results of the experiments described in chapter 4 needs to be evaluated in future studies. A simple way to investigate this possibility would be to test participants with a double-task, with one part congruent and one incongruent to the modality of the learning phase. For example, social participants should be tested not only with a computer task, but also with the same version of the task performed with an experimenter; better results for social participants when learning and retrieving words in the same modality would then suggest a transfer appropriate processing effect.

5.3.2.2 Long-term memory

The investigation of long-term memory traces moves away from the field of learning processes into the broader field of memory research. In fact, the theoretical distinction between learning and memory is a subtle one, as demonstrated by the American Psychological Association's definition: "*Learning and memory are closely related concepts. Learning is the acquisition of skill or knowledge, while memory is the expression of what you've acquired. Another difference is the speed with which the two things happen. If you acquire the new skill or knowledge slowly and laboriously, that's learning. If acquisition occurs instantly, that's making a memory*". Social context has been proven to significantly bias the formation of new memories. In a study by Straube and colleagues (B. Straube et al., 2010), participants watched video clips of an actor speaking directly to them (second-person approach, paragraph 3.1) or to another person (third-person approach), and were thereafter required to recognize the sentences they had just heard and the context in which they were heard. While there was no difference in the recognition of sentences spoken in second- or third-person contexts, source memory (that is, the memory of the context a sentence was heard in) was significantly biased by social interaction, as participants tended to misjudge third-person contexts as second-person contexts. This study corroborates the perspective proposed in previous chapters of this dissertation; it clearly supports the importance of distinguishing between social contexts in which a participant is directly involved in social interaction, and those in which he is merely an observer of social events (Schilbach, 2014; Schilbach et al., 2013). In addition, this evidence provides the first indication that social interaction may significantly modulate the way new memories are formed. However, whether the effects of social interaction extend over time is a question that still remains open.

Generally speaking, the formation of a new memory can be divided into three main stages, namely encoding, consolidation, and retrieval (T. Straube et al., 2008). In the proposed learning game, *encoding* takes place during the learning phase, in which new words are mapped onto their referents. *Retrieval* of the newly learned information occurs during the testing phase. However, as the testing phase takes place immediately after encoding, it does not provide information concerning long-term mnemonic effects. These effects are critically dependent upon *consolidation* processes, during which newly created memory traces are reinforced (Walker & Stickgold, 2004). The efficacy of consolidation depends on several factors, among which sleep seems to play a particularly pivotal role (Atherton et al., 2014; Diekelmann & Born, 2007; Lewis, 2014; Siegel, 2001; Stickgold, 2005). In the task used in the current experiments, the participants' focus was on learning new words; however, the learning context was also manipulated. Future studies are required to evaluate to what extent the manipulation of the context influences the consolidation of memories, by testing retrieval at delayed time points after the learning phase has taken place.

5.4 Clinical application: A therapeutic tool for word re-learning

As proposed in the previous chapters, learning new words is a complex process that many people face even in adulthood. Importantly, this is not only the case for healthy adults acquiring words in a new language, but also in several neurological populations in which the re-learning of words they previously have known is essential for recovery. For example, one of the most common language-related disorders arising as a function of left-hemispheric stroke is the inability to recall words (*anomia*). Depending on the gravity of such anomia and the location and extension of the brain damage, it can become extremely difficult for a patient to communicate and convey meaning, making word re-learning a primary concern in the rehabilitation of these patients.

The method described in this dissertation may provide a new and useful therapeutic tool for patients with language re-learning deficits; indeed, it represents a novel approach, different from those typically used and constituted by long and tedious training sessions, which may be helpful in increasing a patient's motivation. Furthermore, the learning game has the advantage of being short, easy to understand and does not require learning under time pressure. Aphasic patients suffering from anomia and, more specifically, patients with preserved lexical access but impaired word retrieval, represent a particularly suitable target population for this type of training. This patient population is expected to maximally benefit from this approach, as pictures compose the training material and may additionally facilitate word retrieval.

Another clinical application concerns the possibility of using the learning game to differentiate true therapeutic effects from effects due to the presence of a therapist. For example, many language rehabilitation treatments build upon the hypothesis that language homologue areas in the right hemisphere may become involved during recovery of language function[24] (Albert, Sparks, & Helm, 1973; Norton et al., 2009; Schlaug, Marchina, & Norton, 2008). One such treatment is Melodic Intonation Therapy (Norton et al., 2009; S. J. Wilson, Parsons, & Reutens, 2006; Zumbansen, Peretz, & Hébert, 2014), which is a form of music therapy that capitalizes on musical elements of speech, such as melody and rhythm, to improve expressive language by engaging language-homologue regions in the right hemisphere (Norton et al., 2009). In particular, recent evidence strongly suggests that the rhythmical aspects of the therapy (that is, the rhythm of the stimuli and of the therapist's tapping on the patient's hand) are maximally involved in the rehabilitation of speech production (Stahl, Henseler, Turner, Geyer, & Kotz, 2013; Stahl, Kotz, Henseler, Turner, & Geyer, 2011).

An often-neglected aspect in MIT, and many other therapies, is that the presence of a therapist is fundamental. This is particularly problematic as not only do music and social interaction share many common features (see chapters 4.2 and 5.2.1), but also because both aspects through which MIT is supposed to rehabilitate speech (for example, recruitment of right hemispheric regions and the focus on rhythmic aspects), are also important correlates of social interaction (chapter 4). More specifically, the literature reviewed in chapter 1 and the results of the imaging study proposed in chapter 4 provide evidence that social interaction also recruits the language areas of the right hemisphere

---

[24] In case of small left hemispheric lesions, perilesional areas in the left hemisphere are usually recruited to take over the function of the damaged regions; however, in the case of extended lesions, right hemispheric homologues are recruited (for example, Cappa & Vallar, 1992).

(supramarginal and angular gyri, lingual gyrus, and inferior frontal gyrus; chapter 4.5). Furthermore, social interaction induces temporal coordination in virtue of the repetitive temporal behavior emerging between interacting partners (chapter 4.1 and chapter 4.2). Hence, the question arises as to whether effects traditionally attributed to music may, in fact, be dependent upon the presence of a social partner. Study 4 in the current dissertation began to explore this aspect, by comparing the potential effect of music and social interaction as temporally structured oscillators that the learner may synchronize with. While to some extent both music and social interaction lead to an increase in temporal coordination between participants, only the interaction with a partner had an effect on learning behavior. It was concluded that the type of "social rhythm" emerging between a participant and an experimenter may be used to focus a learner's attention, thus facilitating word learning (Louwerse et al., 2012; Rader & Zukow-Goldring, 2012; Rolf et al., 2009). However, is it the same for re-learning? Do patients rely more on the therapist, or are they more focused on music? The learning game introduced here may represent a particularly promising tool for answering this question.

Given the increasing rate of vascular pathologies resulting in speech-related impairments, the lines of research proposed here should be of primary interest to the scientific community. Indeed, the application of the learning game to aphasic patients may significantly contribute to our current understanding of what makes a speech therapy successful; more specifically, the results summarized in this dissertation suggest that part of the positive effect traditionally attributed to a type of therapy may, in fact, be due to the therapist-patient relationship.

5.5 A proposed model for L2 adult word learning

At the beginning of the current dissertation, the question as to how relevant social interaction is in adult word learning was posed. While this question may seem trivial, it reflects a contrast between two opposing positions: On the one hand, adults are considered self-sufficient learners, cognitively equipped to acquire the information they need; for this reason, the presence of other people is not supposed to influence their behavior or their mental processes (Pickering & Garrod, 2004; Stephens et al., 2010). On the other hand, recent evidence suggests that not only is adults' behavior significantly influenced by the presence of others, but also that this influence is qualitatively and quantitatively different when an adult is interacting with someone rather than merely observing them (Ciaramidaro et al., 2014; Schilbach, 2014; Schilbach et al., 2013; Sebastiani et al., 2014). While this position backs up earlier evidence from the field of social psychology (for example, Bond & Titus, 1983; Zajonc, 1965), the implications for adult language learning are far from being understood. Indeed, as seen in the introduction, models of L2 word learning in adults fail to explicitly include social interaction as a factor in this process. For these reasons, a model specifically tailored to adult learners is proposed here. This model attempts to explain and to discuss further the data presented in chapter 4, as well as providing a comprehensive approach for producing testable predictions and newer interpretations of well-known effects such as the role of exposure in second language learning. It must be emphasized that this model focuses on the mapping of a word onto the correct referent; as such, it assumes that the segmentation phase has already been completed (see paragraph 1.2.1 for details).

Evidence collected in the studies summarized in chapter 4 allows us to draw three main conclusions, representing the starting point for the model's evolution: First, social interaction exerts an influence on a learner's behavior, which is reflected in face-to-face settings by the establishment of temporal coordination between learner and partner; second, a social partner directs the learner's attention towards the referent to a new word; and third, local properties of the linguistic context modulate these dynamics, that is, if the linguistic context allows the learner to identify the referent by himself, the presence of another person does not constitute an advantage. In the following paragraphs, these claims will be integrated into a general model for social word learning in healthy adults.

5.5.1 You, me, and the *gavagai*: How a partner may help the learner

Adults and children adopt remarkably similar mechanisms when learning new words (paragraph 1.2.1), and possibly the same neural mechanisms (paragraph 1.2.3). Nevertheless, while most L1 acquisition models explicitly postulate a role for social interaction, L2 learning models only indirectly suggest that *exposure* to a novel language (for example, the context of daily usage) may have some influence on proficiency. However, what exactly may facilitate learning in a social context? Different theories on language acquisition (for example, the social-pragmatic theory, the emergentist coalition model, the social gating hypothesis and, to some extent, the constraints theories; see paragraph 1.2.2) agree on defining a knowledgeable partner as an attentional enhancer and/or modulator. This particular property has been ascribed to a) a caregiver being particularly salient in virtue of her/his characteristics as a complex multi-modal stimulus (Gogate & Bahrick, 2001; Gogate, Walker-Andrews, & Bahrick, 2001), or b) the caregiver attracting attention by interacting with the learner in a contingent and punctual manner compatible with her/his demands (Hollich et al., 2000; Kuhl, 2007; Tomasello, 2000; Tomasello & Carpenter, 2007).

With regard to the first proposition, the multi-modal nature of a social partner certainly favors the learner. In just one exposure, she/he is exposed to several different cues (for example, auditory, visual, somatosensory, linguistic, and meta-linguistic), each of which can be modulated at several levels (for example, for the auditory modality, prosody, content, noise made by movement, interjections, etc.). This abundance of information is difficult to find in other stimuli even when similarly complex (for example, music; see paragraph 4.2; M. Wilson & Wilson, 2005). As redundancy of information coming from several channels has been proven to facilitate allocation of attention (Gogate & Bahrick, 2001; Schmidt-Kassow et al., 2013), focusing on a human partner certainly represents an advantageous choice for the learner.

The second proposition states that a social partner attracts attention because of her/his punctual adaptation to the learner's demands (for example Kuhl, 2007; Pereira et al., 2008; Sage & Baldwin, 2010). This evidence, postulated by L1 acquisition theories, has lately been backed-up by consistent evidence that social partners tend to become temporally coupled (or coordinated) both in their behavior (Demos et al., 2012; Louwerse et al., 2012; Oullier et al., 2008; Pereira et al., 2008; Richardson et al., 2007; Vesper et al., 2011), as well as in their neural activity (Dumas et al., 2010, 2011; Hasson et al., 2004; Jiang et al., 2012; Kawasaki et al., 2013).

At this point, it is vital to note that not only are these two propositions not mutually exclusive, but that a comprehensive account unifying these perspectives may allow us to capture important aspects of word learning. The evidence collected here indeed supports the claim that both attentional mechanisms may be present during social learning (**Figure 48**). As a first step, the presence of a social partner attracts the learner's attention, because it represents an economical way to obtain as much information as possible from a single source; this claim is supported by the results from Study 5 that show how activation in networks and regions implicated in visuo-spatial attention are significantly enhanced in participants who learned socially. This enhancement of attention has two important consequences: First, it allows the creation of a *common ground*, in which high-level cognitive processes (such as understanding the other's intentions, for example Frith & Frith, 2006; paragraph 1.1) take place, allowing for the successful exchange of information. Second, if the learner is paying attention to the knowledgeable partner, her/his reaction to the partner's behavior will be faster (for example Marinovic, Cheung, Riek, & Tresilian, 2014), determining a sort of "*time-locking*" between the learner's response and the partner's action. This result is confirmed by Studies 3 and 4, as well as the training phase of Study 5, which all demonstrated that social participants were significantly faster than participants in the non-social condition. The establishment of a common ground, in which the behavior of the interacting partners is time-locked, ensures that when a referent for a new word is presented, the learner will be maximally ready to grasp its meaning. Associative theories suggest that the proximity in temporal presentation of two stimuli to be mapped, is crucial for successful learning (paragraph 1.2.2.1), and the idea of a *temporal locking* in socially interactive contexts is in line with this theory. However, the creation of a *common ground* is also compatible with the view expressed by social-pragmatic theory, which states that the ability of the learner to understand the other's intentions is even more important (paragraph 1.2.2.3). Certainly, the combination of these two aspects ensures maximum efficiency in the learning process.

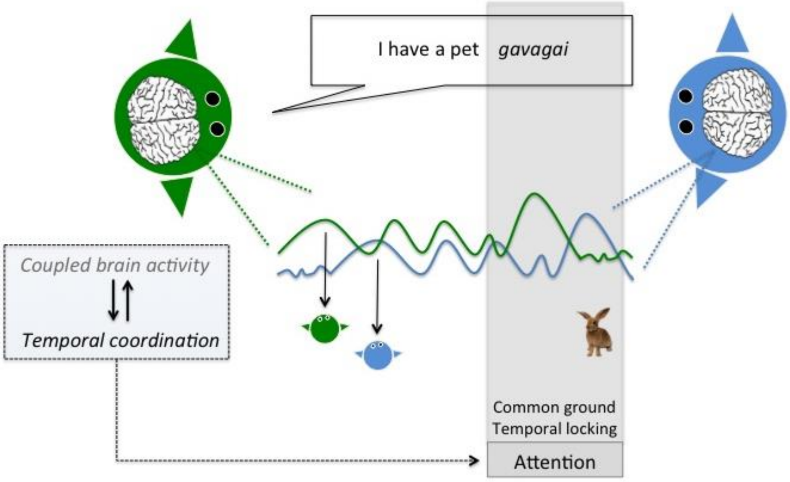
As well as this basic attentional enhancement, spontaneous temporal coordination between interacting partners may additionally enhance the learner's attention. Before explaining why this should be the case, it is crucial to point out that the concepts of temporal coordination and time-locking describe two different phenomena, albeit closely related ones. Time-locking refers to the *temporal proximity* between two events (for example, uttering a novel word and the presentation of its referent), while temporal coordination refers to the *temporal similarity* of two behavioral patterns (for example, similarity in the response of an experimenter and a learner). To better understand the difference between these two phenomena, it is important to consider the pattern of responses during the learning game between the experimenter and the learner. Time-locking predicts that the learner will provide a fast answer *independent of* the speed at which the experimenter is playing the game; temporal coordination, on the other hand, anticipates that the speed of the learner will be *dependent* on the experimenter's speed (and vice versa).

The emergence of temporal coordination between partners may depend on several factors, among them direct visual contact (Oullier, de Guzman, Jantzen, Lagarde, & Kelso, 2008; see paragraph 5.4.1). The effect of temporal coordination is to strengthen the attentional effects of a human partner, by influencing both the percept of a common ground as well as temporal locking. Indeed, temporal coordination often elicits a positive feeling in the people who are interacting by facilitating bonding between partners (Oullier et al., 2008), and also the prediction of a partner's next action (Pecenka et al., 2013; Pecenka & Keller, 2011). It has indeed been explicitly suggested that a reduction in behavioral variability (one of the measures influencing coordination) may represent a strategy through which partners in a joint action task tend to make themselves more predictable, to facilitate coordination (Vesper, Butterfill, Knoblich, & Sebanz, 2010; Vesper et al., 2011, 2012).

The type of interaction between the processes described so far forms a virtual circle of reciprocal influences. Sharing attention creates a common ground which, in turn, facilitates predictions about the other's behavior. These predictions boost temporal locking and temporal coordination, which then also, in turn, reinforce the percept of being on common ground. This would again

Figure 48 – Social word learning model.

Graphical representation of the proposed model. A social partner represents a particularly salient element in the learning environment, enhancing the learner's attention. As a consequence, a common ground is created, in which the learner and partner actions are temporally locked. In face-to-face settings, a behavioral and temporal coupling emerging between the members of the dyad further boosts these processes.



lead to the partner being perceived as "responding to the learner in a punctual and contingent manner" (Kuhl, 2007), boosting her/his attentional potential.

Overall, the final measurable effect is a stable coordinative dynamic, corresponding to the reported effects at both behavioral and neural level[25] (Hasson et al., 2012; Stephens et al., 2010). It has to be noted that this type of interpretation is perfectly in line with the concept of mutual adaptation between interacting people, and also with the behavior of coupled oscillators (paragraph 1.2.2.6 and 5.2).

---

[25] While coupling of neural activities in interacting partners has been consistently reported, it requires an assessment via a hyper-scanning technique. This type of study will certainly be crucial for confirming this model in the future; however, since hyper-scanning was not employed in the work presented here, this aspect (that is, neural coupling) is not explored further here.

5.5.2 The utilitarian learner

Social interaction has been described as an attentional beacon, which can direct the learner's attention through different mechanisms towards relevant elements in the environment. In turn, this is postulated to facilitate learning by creating a common ground, in which the order of the elements to be learned is time-locked and easily predictable. However, is this always the case? The answer to this question decides the position this word-learning model will assume, with respect to competing theories considering the adult learner either as a self-sufficient or a socially grounded entity. The position defended here is a hybrid between these two accounts; while social presence surely influences adult learners, the impact of the exerted influence depends on the task at hand. In particular, the experiments reported here suggest that information from a social partner is particularly useful *when needed*. In the word-learning game, the participants' task was to resolve the indeterminacy of the referent in order to learn new verbal labels. In this context, the idea of the "utilitarian learner" predicts that a social partner's influence will be maximal when the referent is more difficult to find. In other words, the model assumes that when information coming from the sentence context is enough for the learner to ascertain the word meaning, social interaction will not be particularly influential. In the game, a difficult learning situation is represented by new words embedded in a constantly changing sentence context. As a counterpart, new words always repeated as part of the same sentence context represent an easier learning situation, since every time a word is encountered, its referent may immediately be identified on the basis of previous encounters with that word.

In the following paragraphs, the model will be applied to each of these two cases. An example based on Quine's "gavagai" problem (Quine, 1960) is used to introduce the model application in each of the two circumstances. These examples have been proposed in short version in chapter 2. The beginning of the tale is the same for both examples: "*A linguist wanders in a foreign country, but he does not know the local language spoken by the natives. While talking with the linguist about the local habits, a native utters the following sentence:* "*We love animals, and everyone has a pet. For example, I have a* gavagai*". While saying so, the native looks around where several animals are resting on the grass: a rabbit, a cow, and a duck. While with his gesture the native speaker certainly reduced the number of possible referents, there are still three possible meanings for* gavagai.*"*

5.5.2.1 Varied sentence context

Example – […] *Another native intervenes, and says: "I too have a gavagai. My gavagai likes hopping around!". This new information allows the learner to conclude with a reasonable certainty that gavagai means rabbit. Further encounters with the word gavagai may then add information concerning the features that rabbits have, for example being of different colors: "I also have a gavagai; my gavagai is brown, while yours is white".*

This scenario represents an everyday life situation in which a new word is repeatedly encountered, and each time it is embedded in a different sentence context. In the example, several pieces of information concerning the *gavagai* can be collected: It is an animal, it likes hopping, and it comes in different colors (for example, white, brown). How can this learning situation be deconstructed in terms of the model?

To keep it simple, consider the input represented by the new word (*gavagai*), and the possible referents present in the visual scene[26] (**Figure 49**). The definition of the correct referent and its mapping with the novel verbal label

---

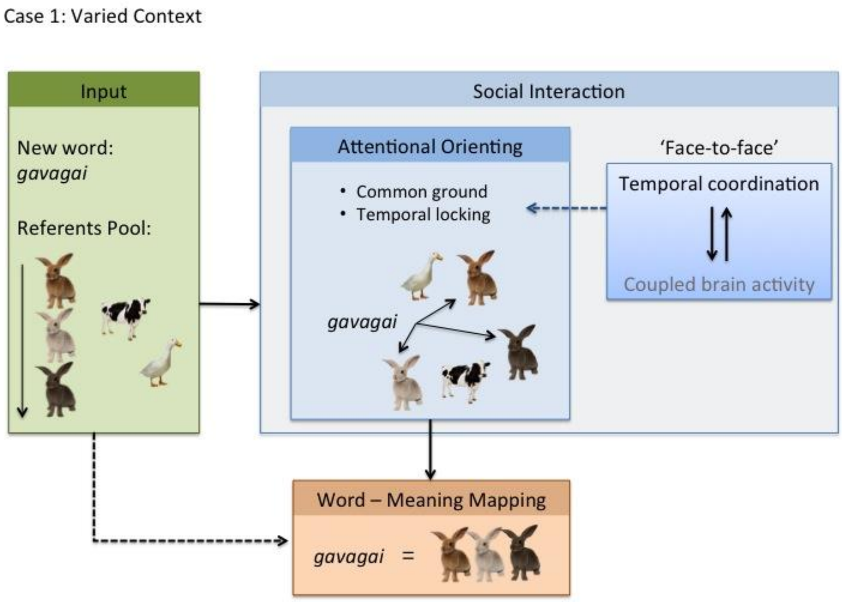
[26] This example refers to a condition in which the referents are visually presented. This situation is not only typical of real-life learning situations, but it was also the situation that participants in the experiments (chapter 4) were exposed to.

may happen via either a “direct” ror an “indirect route”. The *direct route* assumes a direct mapping of the word onto the referent, possibly by guessing strategies based on contextual elements. While this route may be very useful in easier contexts (paragraph 5.3.1.2), in the situation depicted by the example, it would be extremely inefficient for the learner. Indeed, as a minimum, it would require many more examples to collect enough elements to link *gavagai* to rabbit from among the three possible animals. Further, it should be noted that the example depicts a simplified situation in which the choice is restricted to only three possible referents; in real-life situations, the choice is often among infinite referents.

When many referents are encountered, the “indirect” route may be preferred. This route assumes that a social partner facilitates the mapping of a word and a referent. In this scenario, the mechanisms described in paragraph 5.3.2 apply, whereby the presence of a knowledgeable partner enhances the learner’s attention; by creating a common ground in which the event’s order is temporally locked, the partner orients the learner’s attention towards the correct referent and the crucial information in the environment. The possibility of directly observing the social partner and her/his actions (paragraph 5.3.2 and 5.4.1) creates an interactive face-to-face situation, in which the behavior and possibly the neural activities of the interacting partners become temporally coupled. This coupling contributes to the establishment of a common ground and potentiates temporal locking, the combined effect of which is to orient the learner’s attention towards the correct word referent.

As the difficult part of the task (that is, finding the correct referent for a new word in a constantly changing environment) is solved by capitalizing on a knowledgeable partner, the learner can enrich the referent concept with new information added by the context at each occurrence. In the previous example, the final concept of gavagai will be a rich one, containing information about its semantic category (it is an animal), its favorite activities (it likes hopping), and

Figure 49 - Word learning model for varied sentence context ("difficult" learning). When the context of a new word presentation changes at each occurrence, the learner has little information from which to identify the referent from among many possibilities. In this situation, the partner's potential to orient attention may be of significant help in ascertaining the correct referent. As the partner facilitates the identification of the referent, additional information from the context will enrich the representation of the word.



its features (it comes in different colors). This will be of further advantage when generalizing an item to novel contexts, a hypothesis that was confirmed by the results of Study 4 and Study 5. The testing phase of these two experiments required not only to retrieve the meaning of a new word, but also to use new words in novel contexts. In both studies, participants were significantly more accurate in this form of testing when using words, which were learned from varied sentence contexts.

5.5.2.2 Consistent sentence context

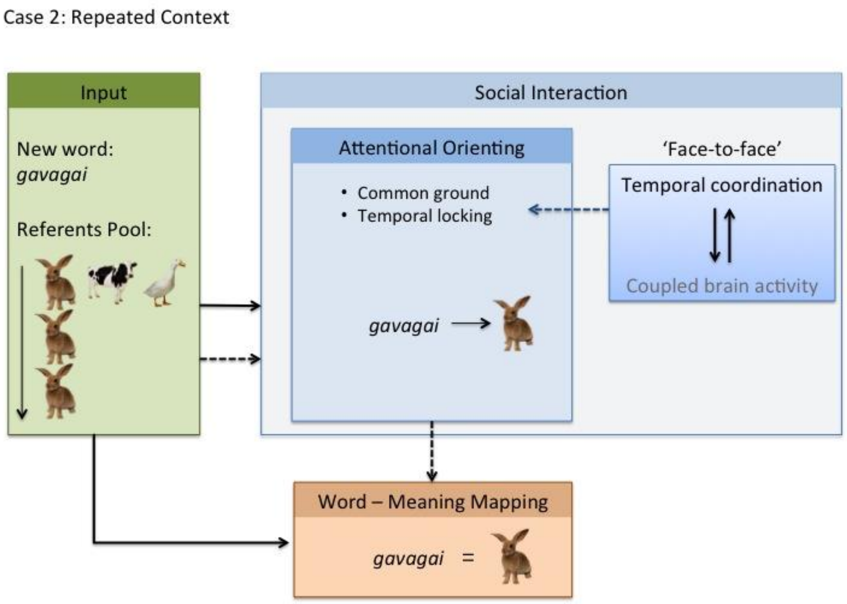
Example – *Another native intervenes, and says: "I too have a gavagai", while still indicating the three animals. However, this time the linguist is more attentive; he notices that the native speaker is referring to the rabbit. Another native jumps into the conversation, but again, he only says: "I too have a gavagai". At this point, however, the linguist has already identified what a gavagai is.*

The example depicts a situation in which the learner is presented with a new word that is always repeated in the same sentence context. In this type of situation, the help that may come from a partner is limited to their first exposures to a novel word. Similar to the example given previously, at the first occurrence of the new word, possible referents include the rabbit, the cow and the duck (**Figure 50**). Unless the context of a presented word strongly suggests a specific meaning, the *indirect route* may be more useful in this initial phase. However, unlike the varied context situation, in the constant scenario the identification of the word meaning reduces the number of possible referents to one – the rabbit. Indeed, the next time *gavagai* is repeated in exactly the same sentence context, the learner will already know what the word means. In this case, the presence of a social partner is not really needed, and a *direct* mapping route can be used. Furthermore, attentional resources initially devoted to the social partner can be re-directed to something else (for example, the referent and its association with the word).

The advantage of this type of learning is that the *direct route* enables faster identification of the referent. Evidence supported by the results reported in chapter 4, for example, show consistently faster response times for words repeated in the same as compared to the different context words. Further, the constant repetition of a word leads to very strong memory traces. However, as a word is always presented within the same sentence context, the number of

Figure 50 - Word learning model for repeated sentence context ("easy" learning).

When the context of a presented new word is the same at each occurrence, the learner can easily recognize the referent from among many possibilities after a first successful identification. In this situation, the partner's potential to orient attention may not be of significant help in ascertaining the correct referent. However, no additional information from the context can be extracted to enrich the representation of the word.



attributes specifying the word meaning is reduced. In the example proposed here, the learner will know that *gavagai* means rabbit, but she/he will not know whether a gavagai can be of different colors or how it moves.

5.5.3 Summary

A model describing word learning in adults has been proposed. More specifically, the model explains how a social partner may help the learner in mapping a new word onto its referent by solving the indeterminacy of the referent problem. This model is based on the data reported in chapter 4, and proposes that social interaction facilitates word learning in healthy adults by working as an attentional spotlight mechanism. This function is supported by several inter-connected mechanisms, such as common ground and temporal locking. These two mechanisms are further reinforced by direct visual contact (face-to-face interaction) with a social partner; indeed, this situation creates an "in person" interaction, in which a behavioral and neural coupling between the partners may take place.

The model assumes an intermediate position between theories of the adult as a self-sufficient versus a social-based learner. Indeed, it is proposed that information coming from a social partner is used only if the learner cannot easily resolve the indeterminacy of the referent by her/himself. When a context provides sufficient information, the learner uses a more efficient and direct route to map a word to its meaning. This route frees up attentional resources otherwise devoted to a social partner, that can then be used to strengthen the association between the word and its meaning.

5.6 Concluding remarks

The current dissertation aimed to investigate the role played by social interaction in word learning in healthy adults. This crucial research question has been highly neglected so far for a number of methodological and theoretical reasons. To overcome these issues, a new method was proposed in chapter 3 to investigate how new words are learned in social compared to non-social contexts. This new method was validated in two preliminary studies, and then employed in a series of three behavioral and fMRI studies, summarized in chapter 4. These experiments demonstrated that social interaction modulates word learning in healthy adults, together with information derived from the local linguistic context. More specifically, a social partner works as an "attentional beacon", able to direct the attention of the learner towards relevant aspects in the environment. This conclusion reveals a similarity between first and second language learning, suggesting that a common learning system may be used in both children and adults (chapter 1). Nevertheless, important differences emerge in the weighting and use of contextual cues, such as those provided by the linguistic context. Furthermore, some shortcoming of the current approach, their possible solutions and possible future directions for research have been outlined, including an extension of the paradigm to clinical populations (for example, aphasic patients).

These results, supported by the literature put forward in the introduction, represent the basis for a new model of word learning in healthy adults, in which two routes of word-meaning mapping have been proposed: A direct “non-social” route, based on contextual linguistic information, and an indirect “social” route, in which a human partner plays a pivotal role in focusing the learner’s attention during learning.

In conclusion, this dissertation significantly adds to our current knowledge of the processes underlying word learning in social interaction, by proposing and validating a new method, by evidencing open questions to address in future investigations, and by delineating a novel model of how social learning processes take place in everyday life.

Bibliography

Abutalebi, J. (2008). Neural aspects of second language representation and language control. *Acta Psychologica*, *128*(3), 466–478.

Abutalebi, J., Annoni, J. M., Zimine, I., Pegna, A. J., Seghier, M. L., Lee-Jahnke, H., … Khateb, A. (2008). Language control and lexical competition in bilinguals: an event-related fMRI study. *Cerebral Cortex*, *18*(7), 1496–1505.

Abutalebi, J., Cappa, S. F., & Perani, D. (2001). The bilingual brain as revealed by functional neuroimaging. *Bilingualism: Language and Cognition*, *4*(02), 179–190. doi:10.1017/S136672890100027X

Abutalebi, J., & Green, D. W. (2008). Control mechanisms in bilingual language production: Neural evidence from language switching studies. *Language and Cognitive Processes*, *23*(4), 557–582.

Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual Diversity, Not Word Frequency, Determines Word-Naming and Lexical Decision Times. *Psychological Science*, *17*(9), 814 –823. doi:10.1111/j.1467-9280.2006.01787.x

Akhtar, N., & Tomasello, M. (1996). Two-year-olds learn words for absent objects and actions. *British Journal of Developmental Psychology*, *14*(1), 79–93.

Akhtar, N., & Tomasello, M. (2000). The social nature of words and word learning. *Becoming a Word Learner: A Debate on Lexical Acquisition*, 115–135.

Albert, M. L., Sparks, R. W., & Helm, N. A. (1973). Melodic intonation therapy for aphasia. *Archives of Neurology*, *29*(2), 130–131.

Allman, M. J. (2011). Deficits in temporal processing associated with autistic disorder. *Frontiers in Integrative Neuroscience*, *5*(2), 1–2.

Allman, M. J., Pelphrey, K. A., & Meck, W. H. (2012). Developmental neuroscience of time and number: implications for autism and other neurodevelopmental disabilities. *Frontiers in Integrative Neuroscience*, *6*.

Allport, A. (1989). Visual attention. Retrieved from http://psycnet.apa.org/psycinfo/1990-97026-016

Allport, F. H. (1924). *Social psychology,*. Boston; New York.

Altenmüller, E., & Schlaug, G. (2013). Neurobiological Aspects of Neurologic Music Therapy. *Music and Medicine*, *5*(4), 210–216.

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nature Reviews Neuroscience*, *7*(4), 268–277.

Anderson, J. R., & Bower, G. H. (1972). Configural properties in sentence memory. *Journal of Verbal Learning and Verbal Behavior*, *11*(5), 594–605. doi:10.1016/S0022-5371(72)80043-4

Anders, S., Heinzle, J., Weiskopf, N., Ethofer, T., & Haynes, J. D. (2011). Flow of affective information between communicating brains. *Neuroimage*, *54*(1), 439–446.

Ardal, S., Donald, M. W., Meuter, R., Muldrew, S., & Luce, M. (1990). Brain responses to semantic incongruity in bilinguals. *Brain and Language*, *39*(2), 187–205.

Astolfi, L., Toppi, J., De Vico Fallani, F., Vecchiato, G., Cincotti, F., Wilke, C., … Babiloni, F. (2011). Imaging the Social Brain by Simultaneous Hyperscanning during Subject Interaction. *IEEE Intelligent Systems*, *26*(5), 38–45.

Astolfi, L., Toppi, J., De Vico Fallani, F., Vecchiato, G., Salinari, S., Mattia, D., … Babiloni, F. (2010). Neuroelectrical hyperscanning measures simultaneous brain activity in humans. *Brain Topography*, *23*(3), 243–256.

Atherton, K. E., Nobre, A. C., Zeman, A. Z., & Butler, C. R. (2014). Sleep-dependent memory consolidation and accelerated forgetting. *Cortex*, *54*, 92–105. doi:10.1016/j.cortex.2014.02.009

Baldwin, D. A. (1993). Early referential understanding: Infants' ability to recognize referential acts for what they are. *Developmental Psychology*, *29*(5), 832.

Baldwin, D. A., Markman, E. M., Bill, B., Desjardins, R. N., Irwin, J. M., & Tidball, G. (1996). Infants' Reliance on a Social Criterion for Establishing Word-Object Relations. *Child Development*, *67*(6), 3135–3153. doi:10.1111/j.1467-8624.1996.tb01906.x

Balsam, P. D., Drew, M. R., & Gallistel, C. R. (2010). Time and Associative

Learning. *Comparative Cognition & Behavior Reviews*, *5*, 1–22.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a “theory of mind”? *Cognition*, *21*(1), 37–46.

Basser, L. S. (1962). Hemiplegia of early onset and the faculty of speech with special reference to the effects of hemispherectomy. *Brain*, *85*(3), 427–460.

Bates, E. (1979). The emergence of symbols: Ontogeny and phylogeny. In *Children's language and communication: The Minnesota Symposium on Child Psychology* (Vol. 12, pp. 121–157).

Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., … Pléh, C. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review*, *10*(2), 344–380.

Batterink, L., & Neville, H. (2013). Implicit and Explicit Second Language Training Recruit Common Neural Mechanisms for Syntactic Processing. *Journal of Cognitive Neuroscience*, *25*(6), 936–951. doi:10.1162/jocn_a_00354

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, *7*(6), 1129–1159.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289–300.

Besson, M., & Kutas, M. (1993). The many facets of repetition: A cued-recall and event-related potential analysis of repeating words in same versus different sentence contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(5), 1115–1133. doi:10.1037/0278-7393.19.5.1115

Bialystok, E., Craik, F. I. M., & Luk, G. (2012). Bilingualism: consequences for mind and brain. *Trends in Cognitive Sciences*, *16*(4), 240–250.

Bialystok, E., & Hakuta, K. (1999). Confounded age: Linguistic and cognitive factors in age differences for second language acquisition. *Second Language Acquisition and the Critical Period Hypothesis*, 161–181.

Bigelow, A. E. (2003). The development of joint attention in blind infants. *Development and Psychopathology*, *15*(02), 259–275. doi:10.1017/S0954579403000142

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where Is the Semantic System? A Critical Review and Meta-Analysis of 120 Functional Neuroimaging Studies. *Cerebral Cortex*, *19*(12), 2767 –2796.

Bird, G., Catmur, C., Silani, G., Frith, C., & Frith, U. (2006). Attention does not modulate neural responses to social stimuli in autism spectrum disorders. *NeuroImage*, *31*(4), 1614–1624. doi:10.1016/j.neuroimage.2006.02.037

Birdsong, D. (1999). *Second Language Acquisition and the Critical Period Hypothesis*. Routledge.

Birdsong, D. (2014). The Critical Period Hypothesis for Second Language Acquisition: Tailoring the Coat of Many Colors. In M. Pawlak & L. Aronin (Eds.), *Essential Topics in Applied Linguistics and Multilingualism* (pp. 43–50). Springer International Publishing.

Birdsong, D., & Molis, M. (2001). On the Evidence for Maturational Constraints in Second-Language Acquisition. *Journal of Memory and Language*, *44*(2), 235–249. doi:10.1006/jmla.2000.2750

Blood, A. J., Zatorre, R. J., Bermudez, P., & Evans, A. C. (1999). Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic brain regions. *Nat Neurosci*, *2*, 382 – 387.

Bloom, K., & Esposito, A. (1975). Social conditioning and its proper control procedures. *Journal of Experimental Child Psychology*, *19*(2), 209–222.

Bloom, K., Russell, A., & Wassenberg, K. (1987). Turn taking affects the quality of infant vocalizations. *Journal of Child Language*, *14*(02), 211–227.

Bloom, L. (1995). *The transition from infancy to language: Acquiring the power of expression*. Cambridge University Press.

Bloom, L. (2000). Pushing the Limits on Theories of Word Learning. *Monographs of the Society for Research in Child Development*, *65*(3), 124–135. doi:10.1111/1540-5834.00100

Bloom, P. (2001). Précis of How children learn the meanings of words. *Behavioral and Brain Sciences*, *24*(06), 1095–1103.

Bloom, P. (2002). Mindreading, communication and the learning of names for things. *Mind & Language*, *17*(1-2), 37–54.

Bloom, P. (2004). Can a Dog Learn a Word? *Science*, *304*(5677), 1605–1606.

doi:10.1126/science.1099899

Boisgueheneuc, F. du, Levy, R., Volle, E., Seassau, M., Duffau, H., Kinkingnehun, S., … Dubois, B. (2006). Functions of the left superior frontal gyrus in humans: a lesion study. *Brain*, *129*(12), 3315–3328. doi:10.1093/brain/awl244

Bond, C. F., & Titus, L. J. (1983). Social facilitation: a meta-analysis of 241 studies. *Psychological Bulletin*, *94*(2), 265.

Bongaerts, T. (1999). Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners. *Second Language Acquisition and the Critical Period Hypothesis*, 133–159.

Bongaerts, T., Mennen, S., & Slik, F. van der. (2000). Authenticity of pronunciation in naturalistic second language acquisition: The case of very advanced late learners of Dutch as a second language. *Studia Linguistica*, *54*(2), 298–308. doi:10.1111/1467-9582.00069

Borovsky, A., Kutas, M., & Elman, J. (2010). Learning to use words: Event-related potentials index single-shot contextual word learning. *Cognition*, *116*(2), 289–296.

Bray, N. (2014). Learning and memory: The hippocampus encodes objects in time. *Nature Reviews Neuroscience*, *15*(5), 281–281. doi:10.1038/nrn3735

Breitenstein, C., Jansen, A., Deppe, M., Foerster, A. F., Sommer, J., Wolbers, T., & Knecht, S. (2005). Hippocampus activity differentiates good from poor learners of a novel lexicon. *Neuroimage*, *25*(3), 958–968.

Breitenstein, C., Kamping, S., Jansen, A., Schomacher, M., & Knecht, S. (2004). Word learning can be achieved without feedback: implications for aphasia therapy. *Restorative Neurology and Neuroscience*, *22*(6), 445–458.

Brent, M. R. (1999). Speech segmentation and word discovery: a computational perspective. *Trends in Cognitive Sciences*, *3*(8), 294–301. doi:10.1016/S1364-6613(99)01350-9

Bressler, S. L., Tang, W., Sylvester, C. M., Shulman, G. L., & Corbetta, M. (2008). Top-down control of human visual cortex by frontal and parietal cortex in anticipatory visual spatial attention. *The Journal of Neuroscience*, *28*(40), 10056–10061.

Brown, T. I., & Stern, C. E. (2014). Contributions of Medial Temporal Lobe and Striatal Memory Systems to Learning and Retrieving Overlapping Spatial Memories. *Cerebral Cortex*, *24*(7), 1906–1922. doi:10.1093/cercor/bht041

Bruner, J. S. (1975). The ontogenesis of speech acts. *Journal of Child Language*, *2*(01), 1–19.

Bruner, J. S. (1983). *Child's Talk.* Norton, New York.

Brysbaert, M., & Duyck, W. (2010). Is it time to leave behind the Revised Hierarchical Model of bilingual language processing after fifteen years of service? *Bilingualism: Language and Cognition*, *13*(03), 359–371. doi:10.1017/S1366728909990344

Buxton, R. B., Uludağ, K., Dubowitz, D. J., & Liu, T. T. (2004). Modeling the hemodynamic response to brain activation. *Neuroimage*, *23*, S220–S233.

Buxton, R. B., Wong, E. C., & Frank, L. R. (1998). Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magnetic Resonance in Medicine*, *39*(6), 855–864.

Calhoun, V. D., Adali, T., McGinty, V. B., Pekar, J. J., Watson, T. D., & Pearlson, G. D. (2001). fMRI activation in a visual-perception task: network of areas detected using the general linear model and independent components analysis. *NeuroImage*, *14*(5), 1080–1088.

Calhoun, V. D., Adali, T., Pearlson, G. D., & Pekar, J. J. (2001). A method for making group inferences from functional MRI data using independent component analysis. *Human Brain Mapping*, *14*(3), 140–151.

Cappa, S. F., & Vallar, G. (1992). The role of the left and right hemispheres in recovery from aphasia. *Aphasiology*, *6*(4), 359–372.

Carey, S. (1978). The Child as Word Learner. In *Linguistic theory and psychological reality* (MIT press.). Cambridge MA: Halle, M., Bresnan, J., Miller, J.A.

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *In Proceedings of the Stanford Child Language Conference*, *15*, 264–278.

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, *63*(4), 1–174.

Carpenter, S. K., Sachs, R. E., Martin, B., Schmidt, K., & Looft, R. (2011).

Learning new vocabulary in German: the effects of inferring word meanings, type of feedback, and time of test. *Psychonomic Bulletin & Review*, *19*, 81–86. doi:10.3758/s13423-011-0185-7

Carr, T. H., Brown, J. S., & Charalambous, A. (1989). Repetition and reading: Perceptual encoding mechanisms are very abstract but not very interactive. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(5), 763–778. doi:10.1037/0278-7393.15.5.763

Carruthers, P. (2002). The cognitive functions of language. *Behavioral and Brain Sciences*, *25*(6), 657–674.

Carter, R. M., Bowling, D. L., Reeck, C., & Huettel, S. A. (2012). A distinct role of the temporal-parietal junction in predicting socially guided decisions. *Science*, *337*(6090), 109–111.

Carter, R. M., & Huettel, S. A. (2013). A nexus model of the temporal–parietal junction. *Trends in Cognitive Sciences*, *17*(7), 328–336.

Chambers, C. D., Payne, J. M., Stokes, M. G., & Mattingley, J. B. (2004). Fast and slow parietal pathways mediate spatial attention. *Nature Neuroscience*, *7*(3), 217–218. doi:10.1038/nn1203

Chee, M. W., Caplan, D., Soon, C. S., Sriram, N., Tan, E. W., Thiel, T., & Weekes, B. (1999). Processing of visually presented sentences in Mandarin and English studied with fMRI. *Neuron*, *23*(1), 127–137.

Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008). Moving on time: brain network for auditory-motor synchronization is modulated by rhythm complexity and musical training. *Journal of Cognitive Neuroscience*, *20*(2), 226–239.

Chomsky, N. A. (1976). *Reflections on language*. New York: Pantheon.

Christophe, A., Dupoux, E., Bertoncini, J., & Mehler, J. (1994). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *The Journal of the Acoustical Society of America*, *95*(3), 1570–1580.

Christophe, A., Gout, A., Peperkamp, S., & Morgan, J. (2003). Discovering words in the continuous speech stream: the role of prosody. *Journal of Phonetics*, *31*(3–4), 585–598. doi:10.1016/S0095-4470(03)00040-8

Chun, M. M., & Jiang, Y. (1998). Contextual Cueing: Implicit Learning and Memory of Visual Context Guides Spatial Attention. *Cognitive Psychology*, *36*(1), 28–71. doi:10.1006/cogp.1998.0681

Church, K. W. (1987). Phonological parsing and lexical retrieval. *Cognition*, *25*(1), 53–69.

Ciaramidaro, A., Becchio, C., Colle, L., Bara, B. G., & Walter, H. (2014). Do you mean me? Communicative intentions recruit the mirror and the mentalizing system. *Social Cognitive and Affective Neuroscience*, *9*(7), 909–916. doi:10.1093/scan/nst062

Clark, E. (1987). The principle of contrast: A constraint on language acquisition. *Mechanisms of Language Acquisition*, 1–33.

Collins, T., Tillmann, B., Barrett, F. S., Delbé, C., & Janata, P. (2014). A combined model of sensory and cognitive representations underlying tonal expectations in music: From audio signals to behavior. *Psychological Review*, *121*(1), 33–65. doi:10.1037/a0034695

Conci, M., & Müller, H. J. (2012). Contextual learning of multiple target locations in visual search. *Visual Cognition*, *20*(7), 746–770. doi:10.1080/13506285.2012.694376

Consonni, M., Cafiero, R., Marin, D., Tettamanti, M., Iadanza, A., Fabbro, F., & Perani, D. (2013). Neural convergence for language comprehension and grammatical class production in highly proficient bilinguals is independent of age of acquisition. *Cortex*, *49*(5), 1252–1258. doi:10.1016/j.cortex.2012.04.009

Constantinidis, C., Bucci, D. J., & Rugg, M. D. (2013). Cognitive functions of the posterior parietal cortex. *Frontiers in Integrative Neuroscience*, *7*. doi:10.3389/fnint.2013.00035

Corbetta, M., Kincade, J. M., Ollinger, J. M., McAvoy, M. P., & Shulman, G. L. (2000). Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nature Neuroscience*, *3*(3), 292–297. doi:10.1038/73009

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The reorienting system of the human brain: from environment to theory of mind. *Neuron*, *58*(3), 306–324.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*(3), 201–215.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684. doi:10.1016/S0022-5371(72)80001-X

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*(3), 268.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, *13*(4), 148–153.

Csibra, G., & Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1567), 1149–1157.

Cui, X., Bryant, D. M., & Reiss, A. L. (2012). NIRS-based hyperscanning reveals increased interpersonal coherence in superior frontal cortex during cooperation. *Neuroimage*, *59*(3), 2430–2437.

Curtiss, S. (1977). *Genie: a psycholinguistic study of a modern-day“ wild child.”* Academic Press New York.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(1), 113–121. doi:10.1037/0096-1523.14.1.113

Decety, J., & Lamm, C. (2007). The role of the right temporoparietal junction in social interaction: how low-level computational processes contribute to meta-cognition. *The Neuroscientist*, *13*(6), 580–593.

De Groot, A. (2006). Effects of stimulus characteristics and background music on foreign language vocabulary learning and forgetting. *Language Learning*, *56*(3), 463–506.

De Jaegher, H., Di Paolo, E., & Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in Cognitive Sciences*, *14*(10), 441–447.

De l’ Etoile, S. K. (2010). Neurologic Music Therapy. *Music and Medicine*, *2*(2), 78.

Demos, A. P., Chaffin, R., Begosh, K. T., Daniels, J. R., & Marsh, K. L. (2012). Rocking to the beat: Effects of music and partner’s movements on spontaneous interpersonal coordination. *Journal of Experimental Psychology: General*, *141*(1), 49.

Dempster, F. N. (1987). Effects of variable encoding and spaced presentations on vocabulary learning. *Journal of Educational Psychology*, *79*(2), 162.

Den Heyer, K. (1986). Manipulating attention-induced priming in a lexical decision task by means of repeated prime-target presentations. *Journal of Memory and Language*, *25*(1), 19–42.

De Ruiter, J. P., Noordzij, M. L., Newman-Norlund, S., Newman-Norlund, R., Hagoort, P., Levinson, S. C., & Toni, I. (2010). Exploring the cognitive infrastructure of communication. *Interaction Studies*, *11*(1), 51–77.

De Ruiter, J. P., Noordzij, M., Newman-Norlund, S., Hagoort, P., & Toni, I. (2007). On the origin of intentions. In *Attention and performance XXII: Sensorimotor foundation of higher cognition*. Oxford: Oxford University press.

De Vignemont, F., & Singer, T. (2006). The empathic brain: how, when and why? *Trends in Cognitive Sciences*, *10*(10), 435–441.

Diekelmann, S., & Born, J. (2007). One memory, two ways to consolidate? *Nature Neuroscience*, *10*(9), 1085–1086. doi:10.1038/nn0907-1085

Diesendruck, G., Markson, L., Akhtar, N., & Reudor, A. (2004). Two-year-olds' sensitivity to speakers' intent: an alternative account of Samuelson and Smith. *Developmental Science*, *7*(1), 33–41. doi:10.1111/j.1467-7687.2004.00320.x

Dijkstra, T., & Van Heuven, W. J. (1998). The BIA model and bilingual word recognition. *Localist Connectionist Approaches to Human Cognition*, 189–225.

Dijkstra, T., & Van Heuven, W. J. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, *5*(03), 175–197.

Dijkstra, T., Van Heuven, W. J., & Grainger, J. (1998). Simulating cross-language competition with the bilingual interactive activation model. *Psychologica Belgica*, *38*(3-4), 177–196.

Dobel, C., Junghöfer, M., Breitenstein, C., Klauke, B., Knecht, S., Pantev, C., & Zwitserlood, P. (2010). New names for known things: on the association of novel word forms with existing semantic information. *Journal of Cognitive Neuroscience*, *22*(6), 1251–1261.

Dominey, P. F., & Dodane, C. (2004). Indeterminacy in language acquisition: the role of child directed speech and joint attention. *Journal of Neurolinguistics*, *17*(2–3), 121–145. doi:10.1016/S0911-6044(03)00056-3

Dufour, R., & Kroll, J. F. (1995). Matching words to concepts in two languages:

A test of the concept mediation model of bilingual representation. *Memory & Cognition*, *23*(2), 166–180. doi:10.3758/BF03197219

Dumas, G., Lachat, F., Martinerie, J., Nadel, J., & George, N. (2011). From social behaviour to brain synchronization: review and perspectives in hyperscanning. *IRBM*, *32*(1), 48–53.

Dumas, G., Nadel, J., Soussignan, R., Martinerie, J., & Garnero, L. (2010). Inter-brain synchronization during social interaction. *PLoS One*, *5*(8), e12166. doi:doi:10.1371/journal.pone.0012166

Eiriksdottir, E., & Catrambone, R. (2011). Procedural Instructions, Principles, and Examples How to Structure Instructions for Procedural Tasks to Enhance Performance, Learning, and Transfer. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *53*(6), 749–770. doi:10.1177/0018720811419154

Elman, J. A., Rosner, Z. A., Cohn-Sheehy, B. I., Cerreta, A. G., & Shimamura, A. P. (2013). Dynamic changes in parietal activation during encoding: Implications for human learning and memory. *NeuroImage*, *82*, 44–52.

Fabbro, F. (2001). The Bilingual Brain: Cerebral Representation of Languages. *Brain and Language*, *79*(2), 211–222. doi:10.1006/brln.2001.2481

Falter, C. M., & Noreika, V. (2011). Interval Timing Deficits and Abnormal Cognitive Development. *Frontiers in Integrative Neuroscience*, *5*(26), 1–2. doi:10.3389/fnint.2011.00026

Falter, C. M., Noreika, V., Wearden, J. H., & Bailey, A. J. (2011). More consistent, yet less sensitive: Interval timing in autism spectrum disorders. *The Quarterly Journal of Experimental Psychology*, 1–15. doi:10.1080/17470218.2012.690770

Ferreri, L., Aucouturier, J.-J., Muthalib, M., Bigand, E., & Bugaiska, A. (2013). Music improves verbal memory encoding while decreasing prefrontal cortex activity: an fNIRS study. *Frontiers in Human Neuroscience*, *7*.

Ferreri, L., Bigand, E., Perrey, S., Muthalib, M., Bard, P., & Bugaiska, A. (2014). Less effort, better results: how does music act on prefrontal cortex in older adults during verbal encoding? An fNIRS study. *Frontiers in Human Neuroscience*, *8*.

Ferreri, L., & Verga, L. (in prep.). Music benefits on verbal learning and memory: When and how does it work?

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age Constraints on

Second-Language Acquisition. *Journal of Memory and Language*, *41*(1), 78–104. doi:10.1006/jmla.1999.2638

Forman, S. D., Cohen, J. D., Fitzgerald, M., Eddy, W. F., Mintun, M. A., & Noll, D. C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magnetic Resonance in Medicine*, *33*(5), 636–647.

Francois, C., & Schön, D. (2010). Learning of musical and linguistic structures: comparing event-related potentials and behavior. *NeuroReport*, *21*(14), 928–932 10.1097/WNR.0b013e32833ddd5e.

François, C., & Schön, D. (2013). Neural sensitivity to statistical regularities as a fundamental biological process that underlies auditory learning: the role of musical practice. *Hearing Research*, *308*, 122–128.

Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *NeuroImage*, *19*(4), 1273–1302. doi:10.1016/S1053-8119(03)00202-7

Frith, C. D., & Frith, U. (1999). Interacting Minds--A Biological Basis. *Science*, *286*(5445), 1692–1695. doi:10.1126/science.286.5445.1692

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, *50*(4), 531–534.

Frith, C. D., & Frith, U. (2012). Mechanisms of social cognition. *Annual Review of Psychology*, *63*, 287–313.

Fromkin, V., Krashen, S., Curtiss, S., Rigler, D., & Rigler, M. (1974). The development of language in Genie: a case of language acquisition beyond the “critical period.” *Brain and Language*, *1*(1), 81–107.

Galantucci, B., & Sebanz, N. (2009). Joint action: current perspectives. *Topics in Cognitive Science*, *1*(2), 255–259.

Gelman, S. A. (2009). Learning from others: children’s construction of concept. *Annual Review of Psychology*, *60*, 115–140. doi:10.1146/annurev.psych.59.103006.093659

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, *73*(2), 135–176. doi:10.1016/S0010-0277(99)00036-0

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*(1), 3–55.

Gleitman, L., Gleitman, H., Landau, B., & Wanner, E. (1988). Where learning begins: Initial representations for language learning. *Language: Psychological and Biological Aspects*, 150–193.

Gleitman, L. R., Newport, E. L., & Gleitman, H. (1984). The current status of the motherese hypothesis. *Journal of Child Language*, *11*(01), 43–79. doi:10.1017/S0305000900005584

Glenberg, A. M. (1976). Monotonic and nonmonotonic lag effects in paired-associate and recognition memory paradigms. *Journal of Verbal Learning and Verbal Behavior*, *15*(1), 1–16.

Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition*, *7*(2), 95–112.

Godden, D. R., & Baddeley, A. D. (1975). Context-Dependent memory in two natural environments: on land and underwater. *British Journal of Psychology*, *66*(3), 325–331. doi:10.1111/j.2044-8295.1975.tb01468.x

Gogate, L. J., & Bahrick, L. E. (2001). Intersensory Redundancy and 7-Month-Old Infants' Memory for Arbitrary Syllable-Object Relations. *Infancy*, *2*(2), 219–231.

Gogate, L. J., Bahrick, L. E., & Watson, J. D. (2000). A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. *Child Development*, *71*(4), 878–894.

Gogate, L. J., Walker-Andrews, A. S., & Bahrick, L. E. (2001). The intersensory origins of word-comprehension: an ecological–dynamic systems view. *Developmental Science*, *4*(1), 1–18.

Golinkoff, R. M., & Hirsh-Pasek, K. (2008). How toddlers begin to learn verbs. *Trends in Cognitive Sciences*, *12*(10), 397–403. doi:10.1016/j.tics.2008.07.003

Golinkoff, R. M., Hirsh-Pasek, K., Bailey, L. M., & Wenger, N. R. (1992). Young children and adults use lexical principles to learn new nouns. *Developmental Psychology*, *28*(1), 99–108. doi:10.1037/0012-1649.28.1.99

Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language*, *14*(01), 23–45.

Graham, S. A., Williams, L. D., & Huber, J. F. (1999). Preschoolers' and Adults' Reliance on Object Shape and Object Function for Lexical Extension. *Journal of Experimental Child Psychology*, *74*(2), 128–151. doi:10.1006/jecp.1999.2514

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1*(02), 67–81.

Green, D. W. (2003). The neural basis of the lexicon and the grammar in L2 acquisition. *The Interface between Syntax and the Lexicon in Second Language Acquisition*, 197–208.

Green, D. W., Crinion, J., & Price, C. J. (2006). Convergence, Degeneracy, and Control. *Language Learning*, *56*, 99–125. doi:10.1111/j.1467-9922.2006.00357.x

Grossmann, T., & Johnson, M. H. (2007). The development of the social brain in human infancy. *European Journal of Neuroscience*, *25*(4), 909–919.

Hakuta, K., Bialystok, E., & Wiley, E. (2003). Critical Evidence A Test of the Critical-Period Hypothesis for Second-Language Acquisition. *Psychological Science*, *14*(1), 31–38. doi:10.1111/1467-9280.01415

Hamilton, A. F. de C. (2013). The mirror neuron system contributes to social responding. *Cortex*, *49*(10), 2957–2959. doi:10.1016/j.cortex.2013.08.012

Han, S., Jiang, Y., Gu, H., Rao, H., Mao, L., Cui, Y., & Zhai, R. (2004). The role of human parietal cortex in attention networks. *Brain*, *127*(3), 650–659.

Hari, R., & Kujala, M. V. (2009). Brain basis of human social interaction: from concepts to brain imaging. *Physiological Reviews*, *89*(2), 453–479.

Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S., & Keysers, C. (2012). Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends in Cognitive Sciences*, *16*(2), 114–121.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science*, *303*(5664), 1634–1640.

Heibeck, T. H., & Markman, E. M. (1987). Word Learning in Children: An Examination of Fast Mapping. *Child Development*, *58*(4), 1021. doi:10.2307/1130543

Heitz, R. P. (2014). The speed-accuracy tradeoff: history, physiology, methodology, and behavior. *Decision Neuroscience*, *8*, 150. doi:10.3389/fnins.2014.00150

Henderson, L., Weighall, A., Brown, H., & Gaskell, G. (2013). Online Lexical Competition During Spoken Word Recognition and Word Learning in Children and Adults. *Child Development*, *84*(5), 1668–1685. doi:10.1111/cdev.12067

Hensch, T. K. (2005). Critical period plasticity in local cortical circuits. *Nature Reviews Neuroscience*, *6*(11), 877–888. doi:10.1038/nrn1787

Heyes, C. M. (1994). Social Learning in Animals: Categories and Mechanisms. *Biological Reviews*, *69*(2), 207–231. doi:10.1111/j.1469-185X.1994.tb01506.x

Hicks, J. L., Marsh, R. L., & Cook, G. I. (2005). An Observation on the Role of Context Variability in Free Recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(5), 1160–1164. doi:10.1037/0278-7393.31.5.1160

Hillecke, T., Nickel, A., & Bolay, H. V. (2005). Scientific perspectives on music therapy. *Annals of the New York Academy of Sciences*, *1060*(1), 271–282.

Hills, T. T., Maouene, J., Riordan, B., & Smith, L. B. (2010). The associative structure of language: Contextual diversity in early word learning. *Journal of Memory and Language*, *63*(3), 259–273.

Hirotani, M., Stets, M., Striano, T., & Friederici, A. D. (2009). Joint attention helps infants learn new words: event-related potential evidence. *NeuroReport*, *20*(6), 600–605.

Hirsh-Pasek, K., & Golinkoff, R. M. (1999). *The origins of grammar: Evidence from early language comprehension*. MIT Press.

Hollich, G. J., Hirsh-Pasek, K., Golinkoff, R. M., Brand, R. J., Brown, E., Chung, H. L., … Bloom, L. (2000). Breaking the Language Barrier: An Emergentist Coalition Model for the Origins of Word Learning. *Monographs of the Society for Research in Child Development*, *65*(3), i–135.

Horwitz, B., Rumsey, J. M., & Donohue, B. C. (1998). Functional connectivity of the angular gyrus in normal reading and dyslexia. *Proceedings of the National Academy of Sciences*, *95*(15), 8939–8944.

Ho, Y. C., Cheung, M. C., & Chan, A. S. (2003). Music training improves verbal but not visual memory: cross-sectional and longitudinal explorations in children. *Neuropsychology*, *17*, 439 – 450.

Huberle, E., & Karnath, H.-O. (2006). Global shape recognition is modulated by the spatial distance of local elements—evidence from simultanagnosia. *Neuropsychologia*, *44*(6), 905–911.

Huettel, S. A., Song, A. W., & McCarthy, G. (2004). *Functional magnetic resonance imaging* (Vol. 1). Sinauer Associates Sunderland, MA.

Hurkmans, J., de Bruijn, M., Boonstra, A. M., Jonkers, R., Bastiaanse, R., Arendzen, H., & Reinders-Messelink, H. A. (2011). Music in the treatment of neurological language and speech disorders: A systematic review. *Aphasiology*, 1–19. doi:10.1080/02687038.2011.602514

Jäncke, L., & Sandmann, P. (2010). Music listening while you learn: No influence of background music on verbal learning. *Behavioral and Brain Functions*, *6*(1), 3. doi:10.1186/1744-9081-6-3

Janik, V. M., & Slater, P. J. B. (2000). The different roles of social learning in vocal communication. *Animal Behaviour*, *60*(1), 1–11. doi:10.1006/anbe.2000.1410

Jeong, H., Hashizume, H., Sugiura, M., Sassa, Y., Yokoyama, S., Shiozaki, S., & Kawashima, R. (2011). Testing Second Language Oral Proficiency in Direct and Semidirect Settings: A Social-Cognitive Neuroscience Perspective. *Language Learning*, *61*(3), 675–699. doi:10.1111/j.1467-9922.2011.00635.x

Jeong, H., Sugiura, M., Sassa, Y., Wakusawa, K., Horie, K., Sato, S., & Kawashima, R. (2010). Learning second language vocabulary: Neural dissociation of situation-based learning and text-based learning. *NeuroImage*, *50*(2), 802–809. doi:10.1016/j.neuroimage.2009.12.038

Jiang, J., Dai, B., Peng, D., Zhu, C., Liu, L. I., Lu, C., & Jing. (2012). Neural synchronization during face-to-face communication. *The Journal of Neuroscience*, *32*(45), 16064–16069. doi:10.1523/JNEUROSCI.2926-12.2012

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*(1), 60–99. doi:10.1016/0010-0285(89)90003-0

Johnson, J. S., & Newport, E. L. (1991). Critical period effects on universal properties of language: The status of subjacency in the acquisition of a second language. *Cognition*, *39*(3), 215–258. doi:10.1016/0010-0277(91)90054-8

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review*, *96*(3), 459.

Jungers, M. K., Palmer, C., & Speer, S. R. (2002). Time after time: The coordinating influence of tempo in music and speech. *Cognitive Processing*, *1*(2), 21–35.

Jusczyk, P. W., Friederici, A. D., Wessels, J. M. I., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants′ Sensitivity to the Sound Patterns of Native Language Words. *Journal of Memory and Language*, *32*(3), 402–420. doi:10.1006/jmla.1993.1022

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The Beginnings of Word Segmentation in English-Learning Infants. *Cognitive Psychology*, *39*(3–4), 159–207. doi:10.1006/cogp.1999.0716

Kahan, J., & Foltynie, T. (2013). Understanding DCM: ten simple rules for the clinician. *Neuroimage*, *83*, 542–549.

Kaminski, J., Call, J., & Fischer, J. (2004). Word Learning in a Domestic Dog: Evidence for “Fast Mapping.” *Science*, *304*(5677), 1682–1683. doi:10.1126/science.1097859

Kang, H. J., & Williamson, V. J. (2013). Background music can aid second language learning. *Psychology of Music*. doi:10.1177/0305735613485152

Kasari, C., Paparella, T., Freeman, S., & Jahromi, L. B. (2008). Language outcome in autism: randomized comparison of joint attention and play interventions. *Journal of Consulting and Clinical Psychology; Journal of Consulting and Clinical Psychology*, *76*(1), 125–137.

Kasari, C., Sigman, M., Mundy, P., & Yirmiya, N. (1990). Affective sharing in the context of joint attention interactions of normal, autistic, and mentally retarded children. *Journal of Autism and Developmental Disorders*, *20*(1), 87–100.

Kawasaki, M., Yamada, Y., Ushiku, Y., Miyauchi, E., & Yamaguchi, Y. (2013). Inter-brain synchronization during coordination of speech rhythm in human-to-human social interaction. *Scientific Reports*, *3*(1692). doi:doi:10.1038/srep01692

Kelso, J. S. (1997). *Dynamic patterns: The self-organization of brain and behavior*. MIT press.

Keysers, C., & Gazzola, V. (2009). Expanding the mirror: vicarious activity for actions, emotions, and sensations. *Current Opinion in Neurobiology*, *19*(6), 666–671.

Kilgour, A. R., Jakobson, L. S., & Cuddy, L. L. (2000). Music training and rate of presentation as mediators of text and song recall. *Memory & Cognition*, *28*(5), 700.

King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R. (2005). Getting to know you: reputation and trust in a two-person economic exchange. *Science*, *308*(5718), 78–83.

Kiviniemi, V., Kantola, J.-H., Jauhiainen, J., Hyvärinen, A., & Tervonen, O. (2003). Independent component analysis of nondeterministic fMRI signal sources. *NeuroImage*, *19*(2), 253–260. doi:10.1016/S1053-8119(03)00097-1

Knoblich, G., & Sebanz, N. (2006). The social nature of perception and action. *Current Directions in Psychological Science*, *15*(3), 99–104.

Knoblich, G., & Sebanz, N. (2008). Evolving intentions for social interaction: from entrainment to joint action. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1499), 2021–2031.

Koffka, K. (1935). *Principles of Gestalt psychology* (Harcourt: Brace and company.). New York: Routledge.

Konvalinka, I., Vuust, P., Roepstorff, A., Frith, C. D., & Ivana. (2010). Follow you, follow me: continuous mutual prediction and adaptation in joint tapping. *The Quarterly Journal of Experimental Psychology*, *63*(11), 2220–2230. doi:10.1080/17470218.2010.497843

Kotz, S. A., D'Ausilio, A., Raettig, T., Begliomini, C., Craighero, L., Fabbri-Destro, M., … Fadiga, L. (2010). Lexicality drives audio-motor transformations in Broca's area. *Brain and Language*, *112*(1), 3–11. doi:10.1016/j.bandl.2009.07.008

Kotz, S. A., & Elston-Güttler, K. (2004). The role of proficiency on processing categorical and associative information in the L2 as revealed by reaction times and event-related brain potentials. *Journal of Neurolinguistics*, *17*(2), 215–235.

Krashen, S. D. (1973). Lateralization, Language Learning, and the Critical Period: Some New Evidence. *Language Learning*, *23*(1), 63–74. doi:10.1111/j.1467-1770.1973.tb00097.x

Kroll, J. F., & Curley, J. (1988). Lexical memory in novice bilinguals: The role of concepts in retrieving second language words. *Practical Aspects of Memory*, *2*(389-395).

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*(2), 149–174.

Kroll, J. F., van Hell, J. G., Tokowicz, N., & Green, D. W. (2010). The Revised Hierarchical Model: A critical review and assessment. *Bilingualism: Language and Cognition*, *13*(03), 373–381. doi:10.1017/S136672891000009X

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience*, *5*(11), 831–843.

Kuhl, P. K. (2007). Is speech learning “gated” by the social brain? *Developmental Science*, *10*(1), 110–120.

Kuhl, P. K. (2010). Brain Mechanisms in Early Language Acquisition. *Neuron*, *67*(5), 713–727.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 979–1000. doi:10.1098/rstb.2007.2154

Kuhl, P. K., Tsao, F. M., & Liu, H. M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences*, *100*(15), 9096–9101. doi:10.1073/pnas.1532872100

Lagarde, J., & Kelso, J. (2006). Binding of movement, sound and touch: multimodal coordination dynamics. *Experimental Brain Research*, *173*(4), 673–688.

Laland, K. N. (2004). Social learning strategies. *Animal Learning & Behavior*, *32*(1), 4–14. doi:10.3758/BF03196002

Landau, B., Smith, L. B., & Jones, S. (1992). Syntactic context and the shape bias in children’s and adults’ lexical learning. *Journal of Memory and Language*, *31*(6), 807–825.

Landau, B., Smith, L., & Jones, S. (1998). Object Shape, Object Function, and

Object Name. *Journal of Memory and Language*, *38*(1), 1–27. doi:10.1006/jmla.1997.2533

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, *106*(1), 119–159. doi:10.1037/0033-295X.106.1.119

Large, E. W., & Palmer, C. (2002). Perceiving temporal regularity in music. *Cognitive Science*, *26*(1), 1–37. doi:10.1016/S0364-0213(01)00057-X

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nat Rev Neurosci*, *9*(12), 920–933.

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: the construct of task-induced involvement. *Applied Linguistics*, *22*(1), 1 –26.

Lemmon, V. W. (1928). The relation of reaction time to measures of intelligence, memory, and learning. *Archives of Psychology*, *94*, 38.

Lenneberg, E. H., Chomsky, N., & Marx, O. (1967). *Biological foundations of language* (Vol. 68). Wiley New York.

Lewis, S. (2014). Sleep and memory: To sleep and learn. *Nature Reviews Neuroscience*, *15*(7), 429–429. doi:10.1038/nrn3782

Lindenberger, U., Li, S. C., Gruber, W., & Müller, V. (2009). Brains swinging in concert: cortical phase synchronization while playing guitar. *BMC Neuroscience*, *10*(1), 10–22.

Liu, T., & Pelowski, M. (2014). Clarifying the interaction types in two-person neuroscience research. *Frontiers in Human Neuroscience*, *8*, 276. doi:10.3389/fnhum.2014.00276

Loehr, J. D., Large, E. W., & Palmer, C. (2011). Temporal coordination and adaptation to rate change in music performance. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(4), 1292.

Lohnas, L. J., Polyn, S. M., & Kahana, M. J. (2011). Contextual variability in free recall. *Journal of Memory and Language*, *64*(3), 249–255. doi:10.1016/j.jml.2010.11.003

Long, M. H. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition*, *12*(03), 251–285.

Louwerse, M. M., Dale, R., Bard, E. G., & Jeuniaux, P. (2012). Behavior

Matching in Multimodal Communication Is Synchronized. *Cognitive Science*, *36*(8), 1404–1426. doi:10.1111/j.1551-6709.2012.01269.x

Ludke, K. M., Ferreira, F., & Overy, K. (2014). Singing can facilitate foreign language learning. *Memory & Cognition*, *42*(1), 41–52.

Macaluso, E., Frith, C. D., & Driver, J. (2000). Modulation of human visual cortex by crossmodal spatial attention. *Science*, *289*(5482), 1206–1208.

Maguire, M. J., Hirsh-Pasek, K., & Golinkoff, R. M. (2006). A Unified Theory of Word Learning: Putting Verb Acquisition in Context. *Action Meets Word: How Children Learn Verbs*, 364.

Maguire, M. J., Hirsh-Pasek, K., Golinkoff, R. M., & Brandone, A. C. (2008). Focusing on the relation: Fewer exemplars facilitate children's initial verb learning and extension. *Developmental Science*, *11*(4), 628–634.

Marinovic, W., Cheung, F., Riek, S., & Tresilian, J. R. (2014). The Effect of Attention on the Release of Anticipatory Timing Actions. *Behavioral Neuroscience*, No Pagination Specified. doi:10.1037/bne0000007

Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science*, *14*(1), 57–77.

Markman, E. M. (1991). The whole-object, taxonomic, and mutual exclusivity assumptions as initial constraints on word meanings. *Perspectives on Language and Thought: Interrelations in Development*, 72–106.

Markman, E. M., & Abelev, M. (2004). Word learning in dogs? *Trends in Cognitive Sciences*, *8*(11), 479–481.

Markman, E. M., & Hutchinson, J. E. (1984). Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive Psychology*, *16*(1), 1–27.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*(2), 121–157.

Markson, L., & Bloom, P. (1997). Evidence against a dedicated system for word learning in children. *Nature*, *385*(6619), 813–815.

Marreiros, A. C., Kiebel, S. J., & Friston, K. J. (2008). Dynamic causal modelling for fMRI: a two-state model. *Neuroimage*, *39*(1), 269–278.

Mather, E., & Plunkett, K. (2012). The Role of Novelty in Early Word

Learning. *Cognitive Science*, *36*(7), 1157–1177. doi:10.1111/j.1551-6709.2012.01239.x

Mathias, B., Palmer, C., Perrin, F., & Tillmann, B. (2014). Sensorimotor Learning Enhances Expectations During Auditory Perception. *Cerebral Cortex*, bhu030. doi:10.1093/cercor/bhu030

Matthews, P. M., & Jezzard, P. (2004). Functional magnetic resonance imaging. *Journal of Neurology, Neurosurgery & Psychiatry*, *75*(1), 6–12.

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, *38*(4), 465–494.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, *88*(5), 375.

McKeown, M. J., Makeig, S., Brown, G. G., Jung, T.-P., Kindermann, S. S., Bell, A. J., & Sejnowski, T. J. (1998). Analysis of fMRI data by blind separation into independent spatial components. *Human Brain Mapping*, *6*, 160–188.

McKeown, M. J., & Sejnowski, T. J. (1998). Independent component analysis of fMRI data: examining the assumptions. *Human Brain Mapping*, *6*(5-6), 368–372.

McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language*, *39*(1), 21–46.

Merriman, W. E., Bowman, L. L., & MacWhinney, B. (1989). The Mutual Exclusivity Bias in Children's Word Learning. *Monographs of the Society for Research in Child Development*, *54*(3/4), i. doi:10.2307/1166130

Mestres-Missé, A. (2007). *Neural correlates of word learning and meaning acquisition*. University of Barcelona, Barcelona.

Mestres-Missé, A., Càmara, E., Rodriguez-Fornells, A., Rotte, M., & Münte, T. F. (2008). Functional neuroanatomy of meaning acquisition from context. *Journal of Cognitive Neuroscience*, *20*(12), 2153–2166.

Mestres-Missé, A., Münte, T. F., & Rodriguez-Fornells, A. (2009). Functional neuroanatomy of contextual acquisition of concrete and abstract words. *Journal of Cognitive Neuroscience*, *21*(11), 2154–2171.

Mestres-Missé, A., Rodriguez-Fornells, A., & Münte, T. F. (2007). Watching the brain during meaning acquisition. *Cerebral Cortex*, *17*(8), 1858–1866.

Mestres-Missé, A., Rodriguez-Fornells, A., & Münte, T. F. (2010). Neural differences in the mapping of verb and noun concepts onto novel words. *NeuroImage*, *49*(3), 2826–2835. doi:10.1016/j.neuroimage.2009.10.018

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*(1), 167–202.

Mol, L., Krahmer, E., Maes, A., Swerts, M., & Lisette. (2012). Adaptation in gesture: Converging hands or converging minds? *Journal of Memory and Language*, *66*(1), 249–264. doi:10.1016/j.jml.2011.07.004

Montague, P. R., Berns, G. S., Cohen, J. D., McClure, S. M., Pagnoni, G., Dhamala, M., … Apple, N. (2002). Hyperscanning: simultaneous fMRI during linked social interactions. *Neuroimage*, *16*(4), 1159–1164.

Moore, C. J., & Price, C. J. (1999). Three Distinct Ventral Occipitotemporal Regions for Reading and Object Naming. *NeuroImage*, *10*(2), 181–192. doi:10.1006/nimg.1999.0450

Moreno, E. M., & Kutas, M. (2005). Processing semantic anomalies in two languages: an electrophysiological exploration in both languages of Spanish–English bilinguals. *Cognitive Brain Research*, *22*(2), 205–220. doi:10.1016/j.cogbrainres.2004.08.010

Mori, S., & Zhang, J. (2006). Principles of diffusion tensor imaging and its applications to basic neuroscience research. *Neuron*, *51*(5), 527–539.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*(5), 519–533. doi:10.1016/S0022-5371(77)80016-9

Moscovitch, M., & Craik, F. I. M. (1976). Depth of processing, retrieval cues, and uniqueness of encoding as factors in recall. *Journal of Verbal Learning and Verbal Behavior*, *15*(4), 447–458. doi:10.1016/S0022-5371(76)90040-2

Moses, L. J., Baldwin, D. A., Rosicky, J. G., & Tidball, G. (2001). Evidence for referential understanding in the emotions domain at twelve and eighteen months. *Child Development*, *72*(3), 718–735.

Mueller, J. L. (2005). Electrophysiological correlates of second language processing. *Second Language Research*, *21*(2), 152–174. doi:10.1191/0267658305sr256oa

Mundy, P. (1995). Joint attention and social-emotional approach behavior in children with autism. *Development and Psychopathology*, *7*(01), 63–82.

Mundy, P., & Jarrold, W. (2010). Infant joint attention, neural networks and social cognition. *Neural Networks*, *23*(8-9), 985–997.

Mundy, P., & Newell, L. (2007). Attention, joint attention, and social cognition. *Current Directions in Psychological Science*, *16*(5), 269–274.

Mundy, P., & Sigman, M. (2006). Joint attention, social competence and developmental psychopathology. *Developmental Psychopathology, Second Edition, Volume One: Theory and Methods, Hoboken, NJ: Wiley*.

Mundy, P., Sigman, M., & Kasari, C. (1990). A longitudinal study of joint attention and language development in autistic children. *Journal of Autism and Developmental Disorders*, *20*(1), 115–128.

Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning Word Meanings From Context During Normal Reading. *American Educational Research Journal*, *24*(2), 237–270.

Nardo, D., Santangelo, V., & Macaluso, E. (2014). Spatial orienting in complex audiovisual environments. *Human Brain Mapping*, *35*(4), 1597–1614. doi:10.1002/hbm.22276

Nelson, K. (1988). Constraints on word learning? *Cognitive Development*, *3*(3), 221–246.

Newman-Norlund, R. D., Noordzij, M. L., Meulenbroek, R., & Bekkering, H. (2007). Exploring the brain basis of joint action: co-ordination of actions, goals and intentions. *Social Neuroscience*, *2*(1), 48–65.

Newman-Norlund, S., Noordzij, M. L., Newman-Norlund, R. D., Volman, I. A. C., Ruiter, J. P., Hagoort, P., & Toni, I. (2009). Recipient design in tacit communication. *Cognition*, *111*(1), 46–54.

Newport, E. L. (1988). Constraints on learning and their role in language acquisition: Studies of the acquisition of American sign language. *Language Sciences*, *10*(1), 147–172. doi:10.1016/0388-0001(88)90010-1

Newport, E. L. (1990). Maturational Constraints on Language Learning. *Cognitive Science*, *14*(1), 11–28. doi:10.1207/s15516709cog1401_2

Noordzij, M. L., Newman-Norlund, S. E., De Ruiter, J. P., Hagoort, P., Levinson, S. C., & Toni, I. (2009). Brain mechanisms underlying human communication. *Frontiers in Human Neuroscience*, *3*, 1–14.

Norris, D., McQueen, J. M., Cutler, A., & Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology*, *34*(3), 191–243.

Norton, A., Zipse, L., Marchina, S., & Schlaug, G. (2009). Melodic intonation therapy: shared insights on how it is done and why it might help. *Annals of the New York Academy of Sciences*, *1169*(1), 431–436.

Nummenmaa, L., & Calder, A. J. (2009). Neural mechanisms of social attention. *Trends in Cognitive Sciences*, *13*(3), 135–143.

Oberman, L. M., & Ramachandran, V. S. (2007). The simulating social mind: the role of the mirror neuron system and simulation in the social and communicative deficits of autism spectrum disorders. *Psychological Bulletin*, *133*(2), 310.

Ochsner, K. N., & Lieberman, M. D. (2001). The emergence of social cognitive neuroscience. *American Psychologist*, *56*(9), 717–734. doi:10.1037/0003-066X.56.9.717

Ogawa, S., & Sung, Y.-W. (2007). Functional magnetic resonance imaging. *Scholarpedia*, *2*(10), 3105.

Ojemann, G. A., & Whitaker, H. A. (1978). The bilingual brain. *Archives of Neurology*, *35*(7), 409.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113.

Opitz, B., & Friederici, A. D. (2004). Brain correlates of language learning: the neuronal dissociation of rule-based versus similarity-based learning. *The Journal of Neuroscience*, *24*(39), 8436–8440.

Oullier, O., de Guzman, G. C., Jantzen, K. J., Lagarde, J., & Kelso, J. A. S. (2008). Social coordination dynamics: Measuring human bonding. *Social Neuroscience*, *3*(2), 178–192.

Paradis, M. (1994). Neurolinguistic aspects of implicit and explicit memory: Implications for bilingualism and SLA. *Implicit and Explicit Learning of Languages*, *393*, 419.

Paradis, M. (1997). The cognitive neuropsychology of bilingualism. In A. M. B. de & J. F. Kroll (Eds.), *Tutotials in bilingualism: Psicolinguistic perspectives* (pp. 331–354). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Paradis, M. (2004). *A neurolinguistic theory of bilingualism* (Vol. 18). John Benjamins Publishing.

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, *42*(1), 107–123. doi:10.1016/S0042-6989(01)00250-4

Pecenka, N., Engel, A., & Keller, P. E. (2013). Neural correlates of auditory temporal predictions during sensorimotor synchronization. *Frontiers in Human Neuroscience*, *7*.

Pecenka, N., & Keller, P. E. (2011). The role of temporal prediction abilities in interpersonal sensorimotor synchronization. *Experimental Brain Research*, *211*(3-4), 505–515. doi:10.1007/s00221-011-2616-0

Penfield, W., & Roberts, L. (1959). *Speech and brain mechanisms.* Princeton, NJ, US: Princeton University Press.

Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J., & Nichols, T. E. (2011). *Statistical Parametric Mapping: The Analysis of Functional Brain Images: The Analysis of Functional Brain Images*. Academic Press.

Perani, D., Abutalebi, J., Paulesu, E., Brambati, S., Scifo, P., Cappa, S. F., & Fazio, F. (2003). The role of age of acquisition and language usage in early, high-proficient bilinguals: An fMRI study during verbal fluency. *Human Brain Mapping*, *19*(3), 170–182. doi:10.1002/hbm.10110

Perani, D., Paulesu, E., Galles, N. S., Dupoux, E., Dehaene, S., Bettinardi, V., … Mehler, J. (1998). The bilingual brain. Proficiency and age of acquisition of the second language. *Brain*, *121*(10), 1841–1852.

Pereira, A. F., Smith, L. B., & Yu, C. (2008). Social coordination in toddler's word learning: interacting systems of perception and action. *Connection Science*, *20*(2-3), 73–89. doi:10.1080/09540090802091891

Perruchet, P., & Vinter, A. (1998). PARSER: A Model for Word Segmentation. *Journal of Memory and Language*, *39*(2), 246–263. doi:10.1006/jmla.1998.2576

Perry, L. K., Samuelson, L. K., Malloy, L. M., & Schiffer, R. N. (2010). Learn Locally, Think Globally. *Psychological Science*, *21*(12), 1894 –1902. doi:10.1177/0956797610389189

Peterson, D. A., & Thaut, M. H. (2007). Music increases frontal EEG coherence during verbal learning. *Neuroscience Letters*, *412*(3), 217–221.

Phillips-Silver, J., & Trainor, L. J. (2005). Feeling the beat: Movement influences infant rhythm perception. *Science*, *308*(5727), 1430–1430.

Pickering, M. J., & Garrod, S. (2004). The interactive-alignment model: Developments and refinements. *Behavioral and Brain Sciences*, *27*(02), 212–225.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(02), 169–190. doi:10.1017/S0140525X04000056

Pinker, S. (2010). The cognitive niche: Coevolution of intelligence, sociality, and language. *Proceedings of the National Academy of Sciences*, *107*(Supplement 2), 8993–8999.

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics*, *29*(2), 191–215. doi:10.1006/jpho.2001.0134

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, *116*(1), 129.

Potter, M. C., So, K.-F., Eckardt, B. V., & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior*, *23*(1), 23–38.

Price, C. J. (1998). The functional anatomy of word comprehension and production. *Trends in Cognitive Sciences*, *2*(8), 281–288. doi:10.1016/S1364-6613(98)01201-7

Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences*, *22*, 253–279.

Quine, W. V. O. (1960). *Word and object*. Cambridge: MIT press.

Rader, N. de V., & Zukow-Goldring, P. (2012). Caregivers' gestures direct infant attention during early word learning: the importance of dynamic synchrony. *Language Sciences*, *34*(5), 559 – 568. doi:http://dx.doi.org/10.1016/j.langsci.2012.03.011

Raichle, MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences*, *98*(2), 676.

Rainey, D. W., & Larsen, J. D. (2002). The effect of familiar melodies on initial learning and long-term memory for unconnected text. *Music Perception*, *20*(2), 173–186.

Redcay, E., Dodell-Feder, D., Pearrow, M. J., Mavros, P. L., Kleiner, M., Gabrieli, J. D. E., & Saxe, R. (2010). Live face-to-face interaction during fMRI: A new tool for social cognitive neuroscience. *Neuroimage*, *50*(4), 1639–1647.

Reed, A. V. (1973). Speed-Accuracy Trade-Off in Recognition Memory. *Science*, *181*(4099), 574–576. doi:10.1126/science.181.4099.574

Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review*, *12*(6), 969–992. doi:10.3758/BF03206433

Repp, B. H., & Keller, P. E. (2008). Sensorimotor synchronization with adaptively timed sequences. *Human Movement Science*, *27*(3), 423–456.

Repp, B. H., & Su, Y.-H. (2013). Sensorimotor synchronization: A review of recent research (2006–2012). *Psychonomic Bulletin & Review*, *20*(3), 403–452. doi:10.3758/s13423-012-0371-2

Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R., & Schmidt, R. C. (2007). Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human Movement Science*, *26*(6), 867–891.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.*, *27*, 169–192.

Rizzolatti, G., & Fabbri-Destro, M. (2008). The mirror system and its role in social cognition. *Current Opinion in Neurobiology*, *18*(2), 179–184.

Rodríguez-Fornells, A., Cunillera, T., Mestres-Missé, A., & de Diego-Balaguer, R. (2009). Neurophysiological mechanisms involved in language learning in adults. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1536), 3711–3735.

Rodriguez-Fornells, A., De Diego Balaguer, R., & Münte, T. F. (2006). Executive control in bilingual language processing. *Language Learning*, *56*(s1), 133–190.

Rolf, M., Hanheide, M., & Rohlfing, K. J. (2009). Attention via synchrony: Making use of multimodal cues in social learning. *Autonomous Mental Development, IEEE Transactions on*, *1*(1), 55–67.

Rossi, V., & Pourtois, G. (2014). Electrical neuroimaging reveals content-specific effects of threat in primary visual cortex and fronto-parietal attentional networks. *NeuroImage*, *98*, 11–22. doi:10.1016/j.neuroimage.2014.04.064

Saalmann, Y. B., Pigarev, I. N., & Vidyasagar, T. R. (2007). Neural mechanisms of visual attention: how top-down feedback highlights relevant locations. *Science*, *316*(5831), 1612–1615.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word Segmentation: The Role of Distributional Cues. *Journal of Memory and Language*, *35*(4), 606–621. doi:10.1006/jmla.1996.0032

Sage, K. D., & Baldwin, D. (2010). Social gating and pedagogy: Mechanisms for learning and implications for robotics. *Neural Networks*, *23*(8-9), 1091–1098. doi:10.1016/j.neunet.2010.09.004

Sage, K. D., & Baldwin, D. (2011). Disentangling the Social and the Pedagogical in Infants' Learning about Tool-Use. *Social Development*, *20*(4), 825–844. doi:10.1111/j.1467-9507.2011.00624.x

Saito, D. N., Tanabe, H. C., Izuma, K., Hayashi, M. J., Morito, Y., Komeda, H., … Fujibayashi, Y. (2010). "Stay Tuned": Inter-Individual Neural Synchronization During Mutual Gaze and Joint Attention. *Frontiers in Integrative Neuroscience*, *4*, 1–10.

Samuelson, L. K., & Smith, L. B. (1998). Memory and Attention Make Smart Word Learning: An Alternative Account of Akhtar, Carpenter, and Tomasello. *Child Development*, *69*(1), 94–104. doi:10.1111/j.1467-8624.1998.tb06136.x

Sandrone, S., Bacigaluppi, M., Galloni, M. R., Cappa, S. F., Moro, A., Catani, M., … Martino, G. (2014). Weighing brain activity with the balance: Angelo Mosso's original manuscripts come to light. *Brain*, *137*(2), 621–633. doi:10.1093/brain/awt091

Sarkamo, T., Tervaniemi, M., Laitinen, S., Forsblom, A., Soinila, S., Mikkonen, M., … Laine, M. (2008). Music listening enhances cognitive recovery and mood after middle cerebral artery stroke. *Brain*, *131*(3), 866.

Saxe, R. (2006). Uniquely human social cognition. *Current Opinion in Neurobiology*, *16*(2), 235–239.

Schellenberg, E. G. (2003). Does exposure to music have beneficial side effects? In *The cognitive neuroscience of music* (pp. 430–448). New York, NY, US: Oxford University Press.

Schilbach, L. (2014). On the relationship of online and offline social cognition. *Frontiers in Human Neuroscience*, *8*, 278. doi:10.3389/fnhum.2014.00278

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, *36*(04), 393–414.

Schippers, M. B., Gazzola, V., Goebel, R., & Keysers, C. (2009). Playing Charades in the fMRI: Are Mirror and/or Mentalizing Areas Involved in Gestural Communication? *PLoS ONE*, *4*(8), e6801. doi:10.1371/journal.pone.0006801

Schippers, M. B., Roebroeck, A., Renken, R., Nanetti, L., & Keysers, C. (2010). Mapping the information flow from one brain to another during gestural communication. *Proceedings of the National Academy of Sciences*, *107*(20), 9388–9393.

Schlaug, G., Marchina, S., & Norton, A. (2008). From singing to speaking: Why singing may lead to recovery of expressive language function in patients with Broca’s aphasia. *Music Perception*, *25*(4), 315.

Schmidt-Kassow, M., Heinemann, L. V., Abel, C., & Kaiser, J. (2013). Auditory–motor synchronization facilitates attention allocation. *NeuroImage*, *82*, 101–106.

Schmidt, R. C., Fitzpatrick, P., Caron, R., & Mergeche, J. (2011). Understanding social motor coordination. *Human Movement Science*, *30*(5), 834–845. doi:10.1016/j.humov.2010.05.014

Schneider, D., Nott, Z. E., & Dux, P. E. (2014). Task instructions and implicit theory of mind. *Cognition*, *133*(1), 43–47. doi:10.1016/j.cognition.2014.05.016

Schön, D., Boyer, M., Moreno, S., Besson, M., Peretz, I., & Kolinsky, R. (2008). Songs as an aid for language acquisition. *Cognition*, *106*(2), 975–983.

Sebastiani, V., de Pasquale, F., Costantini, M., Mantini, D., Pizzella, V., Romani, G. L., & Della Penna, S. (2014). Being an agent or an observer: Different spectral dynamics revealed by MEG. *NeuroImage*, *in press*. doi:10.1016/j.neuroimage.2014.08.031

Shannon, C. E., & Weaver, W. (1963). *Mathematical theory of communication*. University Illinois Press.

Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The Expected Value of Control: An Integrative Theory of Anterior Cingulate Cortex Function. *Neuron*, *79*(2), 217–240. doi:10.1016/j.neuron.2013.07.007

Shockley, K., Baker, A. A., Richardson, M. J., & Fowler, C. A. (2007). Articulatory constraints on interpersonal postural coordination. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(1), 201–208.

Shockley, K., Richardson, D. C., & Dale, R. (2009). Conversation and coordinative structures. *Topics in Cognitive Science*, *1*(2), 305–319.

Sholl, A., Sankaranarayanan, A., & Kroll, J. F. (1995). Transfer between picture naming and translation: A test of asymmetries in bilingual memory. *Psychological Science*, *6*(1), 45–49.

Shukla, M., Nespor, M., & Mehler, J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology*, *54*(1), 1–32. doi:10.1016/j.cogpsych.2006.04.002

Siegel, J. M. (2001). The REM Sleep-Memory Consolidation Hypothesis. *Science*, *294*(5544), 1058–1063. doi:10.1126/science.1063049

Simmons-Stern, N. R., Budson, A. E., & Ally, B. A. (2010). Music as a Memory Enhancer in Patients with Alzheimer’s Disease. *Neuropsychologia*, *48*(10), 3164 – 3167.

Singer, T. (2006). The neuronal basis and ontogeny of empathy and mind reading: Review of literature and implications for future research. *Neuroscience & Biobehavioral Reviews*, *30*(6), 855–863. doi:10.1016/j.neubiorev.2006.06.011

Smith, D. V., Hayden, B. Y., Truong, T.-K., Song, A. W., Platt, M. L., & Huettel, S. A. (2010). Distinct value signals in anterior and posterior ventromedial prefrontal cortex. *The Journal of Neuroscience*, *30*(7), 2490–2495.

Smith, L. B. (2000). Learning how to learn words: An associative crane. *Becoming a Word Learner: A Debate on Lexical Acquisition*, 51–80.

Smith, L. B., Jones, S. S., & Landau, B. (1996). Naming in young children: a dumb attentional mechanism? *Cognition*, *60*(2), 143–171. doi:10.1016/0010-0277(96)00709-3

Stahl, B., Henseler, I., Turner, R., Geyer, S., & Kotz, S. A. (2013). How to engage the right brain hemisphere in aphasics without even singing: evidence for two paths of speech recovery. *Frontiers in Human Neuroscience*, *7*.

Stahl, B., Kotz, S. A., Henseler, I., Turner, R., & Geyer, S. (2011). Rhythm in disguise: why singing may not hold the key to recovery from aphasia. *Brain*, *134*(10), 3083–3093.

Stein, B. S. (1978). Depth of processing reexamined: The effects of the precision of encoding and test appropriateness. *Journal of Verbal Learning and Verbal Behavior*, *17*(2), 165–174. doi:10.1016/S0022-5371(78)90128-7

Stephan, K. E., Harrison, L. M., Kiebel, S. J., David, O., Penny, W. D., & Friston, K. J. (2007). Dynamic causal models of neural system dynamics: current state and future extensions. *Journal of Biosciences*, *32*(1), 129–144.

Stephan, K. E., Penny, W. D., Moran, R. J., den Ouden, H. E., Daunizeau, J., & Friston, K. J. (2010). Ten simple rules for dynamic causal modeling. *Neuroimage*, *49*(4), 3099–3109.

Stephan, K. E., Tittgemeyer, M., Knösche, T. R., Moran, R. J., & Friston, K. J. (2009). Tractography-based priors for dynamic causal models. *Neuroimage*, *47*(4), 1628–1638.

Stephens, G. J., Silbert, L. J., & Hasson, U. (2010). Speaker-listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences*, *107*(32), 14425–14430. doi:10.1073/pnas.1008662107

Steyvers, M., & Malmberg, K. J. (2003). The effect of normative context variability on recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(5), 760–766. doi:10.1037/0278-7393.29.5.760

Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature*, *437*(7063), 1272–1278. doi:10.1038/nature04286

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., … Yoon, K. E. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, *106*(26), 10587–10592.

Straube, B., Green, A., Chatterjee, A., & Kircher, T. (2010). Encoding Social Interactions: The Neural Correlates of True and False Memories. *Journal of Cognitive Neuroscience*, *23*(2), 306–324. doi:10.1162/jocn.2010.21505

Straube, T., Schulz, A., Geipel, K., Mentzel, H. J., & Miltner, W. H. R. (2008). Dissociation between singing and speaking in expressive aphasia: The role of song familiarity. *Neuropsychologia*, *46*(5), 1505–1512.

Strauss, B. (2002). Social facilitation in motor tasks: a review of research and theory. *Psychology of Sport and Exercise*, *3*(3), 237–256. doi:10.1016/S1469-0292(01)00019-X

Strube, M. J. (2005). What did Triplett really find? A contemporary analysis of the first experiment in social psychology. *The American Journal of Psychology*, 271–286.

Swanborn, M. S. L., & De Glopper, K. (1999). Incidental word learning while reading: A meta-analysis. *Review of Educational Research*, *69*(3), 261–285.

Syal, S., & Finlay, B. L. (2011). Thinking outside the cortex: social motivation in the evolution and development of language. *Developmental Science*, *14*(2), 417–430.

Szekely, A., D'Amico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G., … Bates, E. (2003). Timed picture naming: Extended norms and validation against previous studies. *Behavior Research Methods*, *35*(4), 621–633.

Szekely, A., D'Amico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G., … Bates, E. (2005). Timed action and object naming. *Cortex*, *41*(1), 7–26.

Szekely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., … Wicha, N. (2004). A new on-line resource for psycholinguistic studies. *Journal of Memory and Language*, *51*(2), 247–250.

Thaut, M. H. (2010). Neurologic Music Therapy in Cognitive Rehabilitation. *Music Perception*, *27*(4), 281–285. doi:10.1525/mp.2010.27.4.281

Thaut, M. H., Gardiner, J. C., Holmberg, D., Horwitz, J., Kent, L., Andrews, G., … McIntosh, G. R. (2009). Neurologic music therapy improves executive function and emotional adjustment in traumatic brain injury rehabilitation. *Annals of the New York Academy of Sciences*, *1169*(1), 406–416.

Thaut, M. H., Peterson, D. A., & McIntosh, G. C. (2005). Temporal Entrainment of Cognitive Functions. *Annals of the New York Academy of Sciences*, *1060*(The Neurosciences and Music II: From Perception to Performance), 243–254.

Thaut, M. H., Peterson, D. A., Sena, K. M., & McIntosh, G. C. (2008). Musical structure facilitates verbal learning in multiple sclerosis. *Music Perception*, *25*(4), 325–330.

Thaut, M. H., Peterson, D., McIntosh, G. C., & Hoemberg, V. (2014). musical mnemonics aid verbal memory and induce learning related brain plasticity in multiple sclerosis. *Frontiers in Human Neuroscience*, *8*, 395.

Thiel, C. M., Zilles, K., & Fink, G. R. (2004). Cerebral correlates of alerting, orienting and reorienting of visuospatial attention: an event-related fMRI study. *Neuroimage*, *21*(1), 318–328.

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-Directed Speech Facilitates Word Segmentation. *Infancy*, *7*(1), 53–71. doi:10.1207/s15327078in0701_5

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, *39*(4), 706–716. doi:10.1037/0012-1649.39.4.706

Thoma, V., & Henson, R. N. (2011). Object representations in ventral and dorsal visual streams: fMRI repetition effects depend on attention and part–whole configuration. *NeuroImage*, *57*(2), 513–525. doi:10.1016/j.neuroimage.2011.04.035

Thompson, R., Moulin, C., Hayre, S., & Jones, R. (2005). Music enhances category fluency in healthy older adults and Alzheimer’s disease patients. *Experimental Aging Research*, *31*(1), 91–99.

Tillmann, B., Janata, P., & Bharucha, J. J. (2003). Activation of the inferior frontal cortex in musical priming. *Cognitive Brain Research*, *16*(2), 145–161.

Tognoli, E., Lagarde, J., DeGuzman, G. C., & Kelso, J. (2007). The phi complex as a neuromarker of human social coordination. *Proceedings of* *the National Academy of Sciences*, *104*(19), 8190–8195.

Tomasello, M. (1992). The social bases of language acquisition. *Social Development*, *1*(1), 67–87.

Tomasello, M. (1995). Joint attention as social cognition. In *Joint attention: Its origins and role in development* (pp. 103–130).

Tomasello, M. (2000). The Social-Pragmatic Theory Of Word Learning. *Pragmatics*, *10*(4).

Tomasello, M. (2001). Perceiving intentions and learning words in the second year of life. *Language Acquisition and Conceptual Development*, 132–158.

Tomasello, M., & Barton, M. E. (1994). Learning words in nonostensive contexts. *Developmental Psychology*, *30*(5), 639.

Tomasello, M., & Carpenter, M. (2007). Shared intentionality. *Developmental Science*, *10*(1), 121–125.

Triplett, N. (1898). The Dynamogenic Factors in Pacemaking and Competition. *The American Journal of Psychology*, *9*(4), 507–533. doi:10.2307/1412188

Tulving, E. (1979). Relation between encoding specificity and levels of processing. *Levels of Processing in Human Memory*, 405–428.

Uddin, L. Q., Iacoboni, M., Lange, C., & Keenan, J. P. (2007). The self and social cognition: the role of cortical midline structures and mirror neurons. *Trends in Cognitive Sciences*, *11*(4), 153–157.

Ullman, M. T. (2001). The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research*, *30*(1), 37–69.

Ullman, M. T. (2005). A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. *Mind and Context in Adult Second Language Acquisition*, 141–178.

Van Heuven, W. J. B., Dijkstra, T., & Grainger, J. (1998). Orthographic Neighborhood Effects in Bilingual Word Recognition. *Journal of Memory and Language*, *39*(3), 458–483. doi:10.1006/jmla.1998.2584

Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *Neuroimage*, *48*(3), 564–584.

Verga, L., Bigand, E., & Kotz, S. A. (in prep.). Play along: Temporal coordination to music and social interaction in second language vocabulary learning.

Verga, L., & Kotz, S. A. (in prep.). Learning together or learning alone? Social interaction in adult word learning.

Verga, L., & Kotz, S. A. (2013). How relevant is social interaction in second language learning? *Frontiers in Human Neuroscience*, *7*(550). doi:10.3389/fnhum.2013.00550

Verghese, A., Kolbe, S. C., Anderson, A. J., Egan, G. F., & Vidyasagar, T. R. (2014). Functional size of human visual area V1: A neural correlate of top–down attention. *NeuroImage*, *93*, 47–52.

Verkoeijen, P. P., Rikers, R. M., & Schmidt, H. G. (2004). Detrimental influence of contextual change on spacing effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(4), 796.

Vesper, C., Butterfill, S., Knoblich, G., & Sebanz, N. (2010). A minimal architecture for joint action. *Neural Networks*, *23*(8–9), 998–1003. doi:10.1016/j.neunet.2010.06.002

Vesper, C., van der Wel, R. P. R. D., Knoblich, G., & Sebanz, N. (2011). Making oneself predictable: reduced temporal variability facilitates joint action coordination. *Experimental Brain Research*, *211*(3), 517–530.

Vesper, C., van der Wel, R. P. R. D., Knoblich, G., & Sebanz, N. (2012). Are You Ready to Jump? Predictive Mechanisms in Interpersonal Coordination. *Experimental Brain Research*, *211*, 517–530.

Vrtička, P., Andersson, F., Grandjean, D., Sander, D., & Vuilleumier, P. (2008). Individual attachment style modulates human amygdala and striatum activation during social appraisal. *PLoS One*, *3*(8), e2868.

Walker, M. P., & Stickgold, R. (2004). Sleep-Dependent Learning and Memory Consolidation. *Neuron*, *44*(1), 121–133. doi:10.1016/j.neuron.2004.08.031

Wanner, E., & Gleitman, L. (1982). *Language acquisition: The state of the art*. CUP Archive.

Watzlawick, P., Bavelas, J. B., & Jackson, D. D. (1967). *Pragmatics of human communication: A study of interactional patterns, pathologies, and paradoxes* (W.W. Norton & Company.). New York.

Waxman, S. R., & Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. *Trends in Cognitive Sciences*, *13*(6), 258–263.

Waxman, S. R., & Kosowski, T. D. (1990). Nouns Mark Category Relations: Toddlers' and Preschoolers' Word-Learning Biases. *Child Development*, *61*(5), 1461–1473. doi:10.1111/j.1467-8624.1990.tb02875.x

Weber-Fox, C., & Neville, H. (1996). Maturational Constraints on Functional Specializations for Language Processing: ERP and Behavioral Evidence in Bilingual Speakers. *Journal of Cognitive Neuroscience*, *8*(3), 231–256. doi:10.1162/jocn.1996.8.3.231

Whiting, E., Chenery, H., Chalk, J., Darnell, R., & Copland, D. (2007). Dexamphetamine enhances explicit new word learning for novel objects. *International Journal of Neuro-Psychopharmacology*, *10*(6), 805–816.

Whiting, E., Chenery, H. J., Chalk, J., Darnell, R., & Copland, D. A. (2008). The explicit learning of new names for known objects is improved by dexamphetamine. *Brain and Language*, *104*(3), 254–261.

Willems, R. M., Benn, Y., Hagoort, P., Toni, I., & Varley, R. (2011). Communicating without a functioning language system: Implications for the role of language in mentalizing. *Neuropsychologia*, *49*(11), 3130–3135.

Willems, R. M., de Boer, M., de Ruiter, J. P., Noordzij, M. L., Hagoort, P., & Toni, I. (2010). A dissociation between linguistic and communicative abilities in the human brain. *Psychological Science*, *21*(1), 8–14.

Willems, R. M., & Varley, R. (2010). Neural insights into the relation between language and communication. *Frontiers in Human Neuroscience*, *4*, 1–8.

Williams, J. H. G., Waiter, G. D., Perra, O., Perrett, D. I., & Whiten, A. (2005). An fMRI study of joint attention experience. *Neuroimage*, *25*(1), 133–140.

Wilson, M., & Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychonomic Bulletin & Review*, *12*(6), 957–968.

Wilson, S. J., Parsons, K., & Reutens, D. C. (2006). Preserved singing in aphasia: A case study of the efficacy of melodic intonation therapy. *Music Perception*, 23–35.

Ye, Z., Mestres-Missé, A., Rodriguez-Fornells, A., & Münte, T. F. (2011). Two distinct neural networks support the mapping of meaning to a novel word. *Human Brain Mapping*, *32*(7), 1081–1090.

Young, D. R., & Bellezza, F. S. (1982). Encoding variability, memory organization, and the repetition effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*(6), 545.

Yu, C., & Ballard, D. H. (2007). A unified model of early word learning: Integrating statistical and social cues. *Neurocomputing*, *70*, 2149–2165.

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, *18*(5), 414–420.

Yu, C., Smith, L. B., Klein, K., & Shiffrin, R. M. (2007). Hypothesis testing and associative learning in cross-situational word learning: Are they one and the same. In *Proceedings of the 29th Annual Conference of the Cognitive Science Society* (pp. 737–742). Citeseer.

Yun, K., Watanabe, K., & Shimojo, S. (2012). Interpersonal body and neural synchronization as a marker of implicit social interaction. *Scientific Reports*, *2*, 959. doi:10.1038/srep00959

Zajonc, F. (1965). Social Facilitation. *Science*, *149*(3681), 269–274.

Zhang, H., Long, Z., Ge, R., Xu, L., Jin, Z., Yao, L., & Liu, Y. (2014). Motor Imagery Learning Modulates Functional Connectivity of Multiple Brain Systems in Resting State. *PLoS ONE*, *9*(1), e85489. doi:10.1371/journal.pone.0085489

Zumbansen, A., Peretz, I., & Hébert, S. (2014). Melodic Intonation Therapy: Back to Basics for Future Research. *Name: Frontiers in Neurology*, *5*(7).

Appendix A: Social Questionnaire

FRAGEBOGEN ZUM EXPERIMENT

Bitte beantworte die folgenden Fragen. Es steht Dir dabei immer eine Skala von 1 bis 5 zur Verfügung.

1. Wie gut hat Dir das Experiment gefallen?

1 überhaupt nicht gut	5 sehr gut

2. Wie zufrieden warst Du mit dem Verhalten Deines Partners?

1 überhaupt nicht zufrieden	5 sehr zufrieden

3. Glaubst Du, die Aufgabe war durch die Anwesenheit Deines Partners einfacher?

1 überhaupt nicht einfacher	5 viel einfacher

4. Glaubst Du, Du hättest es bevorzugt, das Spiel alleine zu spielen?

1 überhaupt nicht bevorzugt	5 sehr bevorzugt

5. Sind Dir Fehler in den Satzanfängen Deines Partners aufgefallen?

1 gar keine Fehler	5 sehr viele Fehler

6. Hat sich Dein Partner Deinem Tempo angepasst?

1 überhaupt nicht angepasst	5 stark angepasst

7. War das Tempo Deines Partners zu schnell?

1 überhaupt nicht zu schell	5 viel zu schell

8. War das Tempo Deines Partners zu langsam?

1 überhaupt nicht zu langsam	5 viel zu langsam

9. Hast Du eine Veränderung (Geschwindigkeit, mehr / weniger Fehler ...) im Verhalten Deines Partners im Laufe des Experiments bemerkt?

☐ Ja Welche? ____________________

☐ Nein

10. Hast Du allgemeine Anmerkungen?

11. Hattest Du den Eindruck, dass Dein Partner eine andere Person war?

1 nie	5 immer

12. Hattest Du den Eindruck, dass Dein Partner keine andere Person (sondern z.B. ein Computer) war?

Vielen Dank!

Appendix B: Non-Social Questionnaire

FRAGEBOGEN ZUM EXPERIMENT

Bitte beantworte die folgenden Fragen. Es steht Dir dabei immer eine Skala von 1 bis 5 zur Verfügung.

1. Wie gut hat Dir das Experiment gefallen?

1	5
überhaupt nicht gut	sehr gut

2. Wie zufrieden warst Du mit den Einstellungen des Programms?

1	5
überhaupt nicht zufrieden	sehr zufrieden

3. Glaubst Du, es wäre einfacher gewesen, die Aufgabe mit einem Partner auszuführen?

1	5
überhaupt nicht einfacher	viel einfacher

4. Glaubst Du, Du hättest es bevorzugt, die Aufgabe mit einem Partner auszuführen?

1	5
überhaupt nicht bevorzugt	völlig bevorzugt

5. Sind Dir Fehler in den Satzanfängen, die der Computer Dir gegeben hat, aufgefallen?

1 überhaupt keine Fehler — 5 sehr viele Fehler

6. Hat sich der Computer Deinem Tempo angepasst?

1 überhaupt nicht angepasst — 5 völlig angepasst

7. War das Tempo des Computers zu schnell?

1 überhaupt nicht zu schell — 5 viel zu schell

8. War das Tempo des Computers zu langsam?

1 überhaupt nicht zu langsam — 5 viel zu langsam

9. Hast Du eine Veränderung im Experiment (Geschwindigkeit, mehr / weniger Fehler …) bemerkt?

☐ Ja Welche?

☐ Nein

10. Hast du allgemeine Anmerkungen?

11. Hattest Du den Eindruck, dass die Anfänge des Satzes von einem Computer vorgegeben wurde?

1 nie — 5 immer

12. Hattest Du den Eindruck, dass Dir die Satzanfänge nicht von einem Computer (z.B. sondern von einer anderen Person) vorgegeben wurden?

1 nie — 5 immer

Vielen Dank!

Appendix C: Questionnaire results

Table 6 - fMRI questionnaire results.

The table summarizes the results of the behavioral questionnaire conducted after the experiment to evaluate the extent to which participants in the social group believed the cover story.

Question	Mean score	Social Group	Non-Social Group
1. Did you enjoy the experiment?	3.82 ± 0.68	3.98 ± 0.65	3.39 ± 0.60
2. How satisfied were you with the behavior of your partner /computer program?	3.80 ± 0.94	4.16 ± 0.79	3.00 ± 0.83
3. How much do you think the presence of another person facilitated/ would have facilitated the task?	2.85 ± 1.28	3.14 ± 1.12	2.06 ± 1.42
4. Would it have been better, in your opinion, to perform the task with a partner / computer?	2.24 ± 0.87	2.42 ± 0.87	1.72 ± 1.15
5. Did your partner/ computer make mistakes during the task?	1.87 ± 1.01	2.04 ± 1.04	1.39 ± 0.78
6. Did the partner/computer adapted to your speed?	2.97 ± 1.25	3.38 ± 1.11	1.83 ± 0.87
7. Was the partner / computer too fast?	2.25 ± 1.12	1.92 ± 1.07	3.17 ± 0.66
8. Was the partner / computer too slow?	2.19 ± 1.11	2.26 ± 1.22	2.00 ± 1.22
11. Did you think you were playing with a human partner / computer?	2.90 ± 1.28	2.52 ± 1.20	3.94 ± 0.88
12. Did you think the other partner was not a human / a computer?	2.71 ± 1.15	2.92 ± 1.12	2.11 ± 1.05

Appendix D: Additional fMRI Results

1. Learning effects

The contrast **checkerboard [beginning > end]** elicited a significantly stronger haemodynamic response in an extended set of areas including, in the right hemisphere, the middle cingulate gyrus, inferior parietal gyrus, superior temporal gyrus and the insula. In the left hemisphere, significant clusters of activation were found in the superior medial gyrus, anterior cingulate gyrus, angular gyrus, insula, and in the superior and middle temporal lobe (**Figure 51a** and **Table 7a).**

The opposite contrast **checkerboard [end > beginning]** showed activations bilaterally in visual areas, including the left inferior occipital gyrus and the right lingual and calcarine gyrus (**Figure 51b** and **Table 7b**).

During the subject extraction, activations for the contrast **subject [beginning > end]** were found in a left lateralized network encompassing the middle temporal gyrus, inferior frontal gyrus, angular gyrus and the middle orbital gyrus. Results are summarized in **Figure 52a** and **Table 8a**.

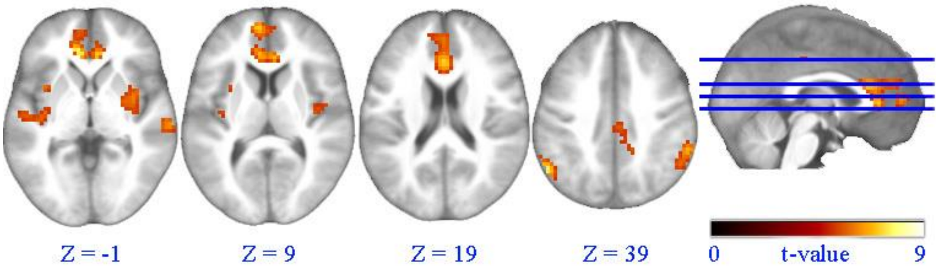
The opposite contrast **subject [end > beginning]** elicited significant activations in a bilateral visual set of areas including the left superior occipital gyrus and the right calcarine gyrus (**Figure 52b** and **Table 8b**).

In the object encoding phase, the contrast **object [beginning > end]** elicited a significant cluster of activation in the left hyppocampus (**Figure 53a** and **Table 9a**).

The contrast **object [end > beginning]** elicited the activation of the thalamus bilaterally, of the left middle occipital gyrus and the right calcarine gyrus. Results are summarized in **Figure 53b** and **Table 9b**.

Figure 51 - Checkerboard: Learning effects (p < .05, FWE-corrected for multiple comparisons at the peak level, k > 30).

51a) Contrast: checkerboard [beginning > end]. The image shows brain areas that are more active at the beginning compared to the end of the learning phase for the observation of the checkerboard. Activations are displayed on axial slices of the average anatomical image of all participants warped to the MNI system of coordinates.



51b) Contrast: checkerboard [end > beginning]. The image shows brain areas that are more active at the end compared to the beginning of the learning phase for the observation of the checkerboard. Activations are displayed on axial slices of the average anatomical image of all participants warped to the MNI system of coordinates.

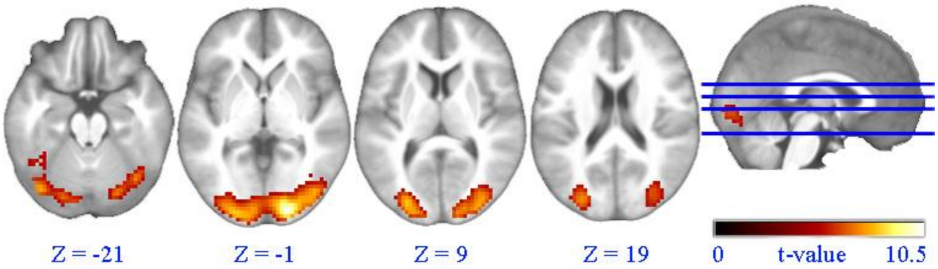


Table 7 - Checkerboard: Learning effects (p < .05, FWE-corrected for multiple comparisons at the peak level, k > 30).

7a) Contrast: checkerboard [beginning > end]. Activations are grouped according to activation clusters; only local maxima are reported.

	Cluster level		Voxel Level	MNI coordinates		
Anatomical Location	**K**	**p-value**	**z-score**	**X**	**Y**	**Z**
Right						
R Middle Cingulate Cx	40	.000	4.85	15	-46	37
			4.81	3	-25	40
R Superior Temporal Gyrus	82	.000	6.20	63	-22	1
R SupraMarginal Gyrus	97	.000	5.48	60	-43	40
R Angular Gyrus			4.98	54	-52	34
R Insula	90	.000	5.18	36	2	-5
				42	10	13
Left						
L Superior Medial Gyrus	455	.000	6.09	-9	56	7
L Anterior Cingulate Cortex				0	29	19
L Angular Gyrus	161	.000	6.61	-60	-58	28
L Insula Lobe	46	.000	5.77	-36	11	-8
L Inferior Temporal Gyrus	72	.000	5.15	-51	-10	29
L Middle Temporal gyrus			5.07	-48	2	-29
L Middle Frontal Gyrus	31	.000	5.22	-18	47	25

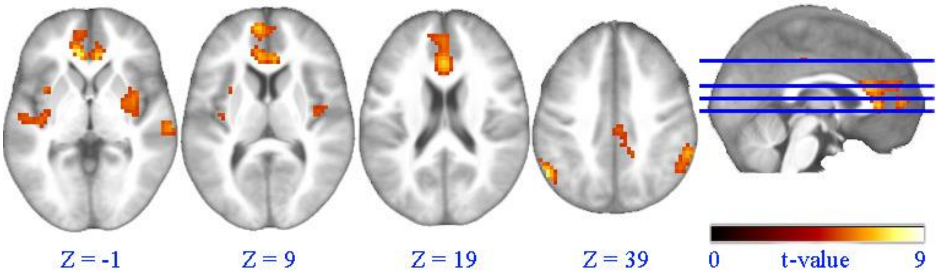
L Superior Temporal Gyrus	91	.000	5.05	-39	-16	-5
L Middle Temporal Gyrus			4.95	-57	-16	-5

7b) Contrast: checkerboard [end > beginning]. Activations are grouped according to activation clusters; only local maxima are reported.

	Cluster level		Voxel Level	MNI coordinates		
Anatomical Location	**K**	**p-value**	**z-score**	**X**	**Y**	**Z**
Right						
R Lingual Gyrus	2759	.000	7.25	18	-91	-5
R Calcarine Gyrus			6.94	-21	-97	-8
L Inferior Occipital Gyrus			6.73	-42	-76	-14

Figure 52 - Subject Extraction: Learning Effects ($p < .05$, FWE-corrected for multiple comparisons, $k > 30$).

52a) Contrast subject [beginning > end]. The image represents areas showing greater activation at the beginning of the learning phase compared to the end. The activations are displayed on axial slices of the average anatomical image of all participants warped to the MNI system of coordinates



52b) Contrast subject [end > beginning]. The image represents areas showing greater activation at the end of the learning phase. The activations are displayed on axial slices of the average anatomical image of all participants warped to the MNI system of coordinates.

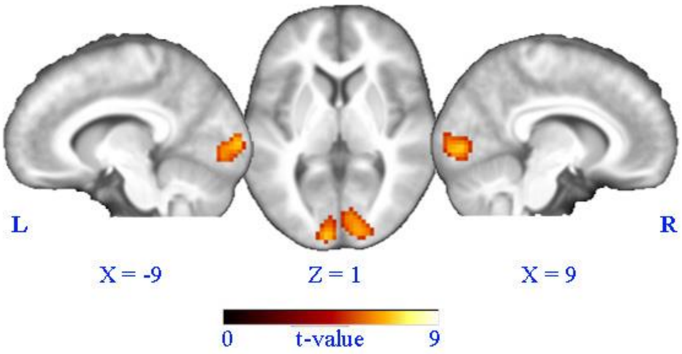


Table 8 - Subject Extraction: Learning Effects (p < .05, FWE-corrected for multiple comparisons, k>30).

8a) Contrast subject [beginning > end]. Specific activations for the contrast **subject** [beginning > end]. Activations are grouped according to activation clusters; only local maxima are reported.

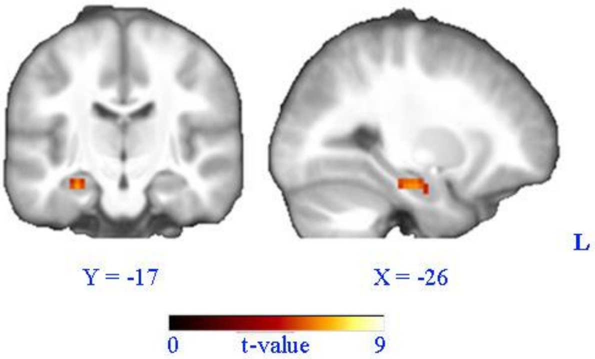
	Cluster level		Voxel Level	MNI coordinates		
Anatomical Location	**K**	**p-value**	**z-score**	**X**	**Y**	**Z**
Left						
L Middle Temporal Gyrus	471	.000	6.62	-60	-13	-20
			6.41	-63	-31	-2
			6.31	-54	-37	-2
L Inferior Frontal Gyrus (p. Orbitalis)	153	.000	6.48	-45	29	-11
L Inferior Frontal Gyrus (p. Triangularis)			5.20	-51	23	13
L Angular Gyrus	74	.000	5.25	-51	-67	25
L Mid Orbital Gyrus	48	.000	5.00	-3	50	-14

8b) Contrast subject [end > beginning]. Specific activations for the contrast **subject** [end > beginning]. Activations are grouped according to activation clusters; only local maxima are reported.

	Cluster level		Voxel Level	MNI coordinates		
Anatomical Location	**K**	**p-value**	**z-score**	**X**	**Y**	**Z**
R Calcarine Gyrus	346	.000	5.74	9	-82	1
L Superior Occipital Gyrus			5.70	-9	-97	4

Figure 53 - Object encoding: learning effects (p < .05, FWE, k > 20).

53a) Contrast object [beginning > end]. Activations are displayed on multiple slices of the average anatomical image of all participants warped to the MNI system of coordinates.



53b) Contrast object [end > beginning]. The activations are displayed on multiple slices of the average anatomical image of all participants warped to the MNI system of coordinates.

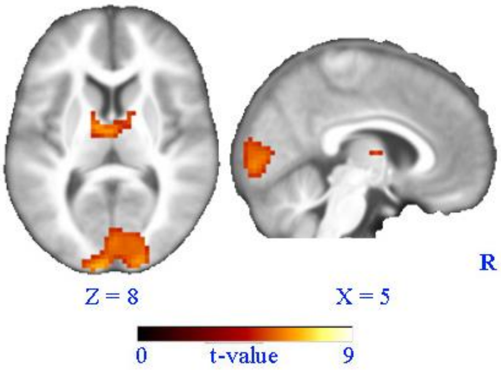


Table 9 - Object Encoding: learning effects ($p < 0.05$, FWE-corrected for multiple comparisons, $k > 20$)

Table 9a - Contrast object [beginning > end]. Specific activations for the contrast **object** [beginning > end]. Activations are grouped according to activation clusters; only local maxima are reported.

	Cluster level		Voxel Level	MNI coordinates		
Anatomical Location	**K**	**P-value**	**z-score**	**X**	**Y**	**Z**
L Hippocampus	22	.004	5.27	-27	-16	-20

Table 9b - Contrast object [end > beginning]. Specific activations for the contrast **object** [end > beginning]. Activations are grouped according to activation clusters; only local maxima are reported.

	Cluster level		Voxel Level	MNI coordinates		
Anatomical Location	**K**	**p-value**	**z-score**	**X**	**Y**	**Z**
L Middle Occipital Gyrus	587	.000	5.87	-9	-100	1
R Calcarine Gyrus			5.41	18	-97	4
L Thalamus (pre-frontal)	287	.000	5.58	-18	-13	16
R Thalamus (pre-frontal)			5.42	18	-10	16

2. Other results

Overall results: At the conservative threshold of p < .05, FWE (cluster level), the activity in the left anterior cingulate cortex druing the checkerboard observation was significantly correlated with the **lag-0 cross-correlation** index, as summarized in **Figure 54** and **Table 10**.

At the liberal threshold of p < .001 uncorrected, a significant positive correlation was found between **test scores** and activity in the bilateral fusiform gyri during the checkerboard observation (**Figure 55** and **Table 11**).

At the liberal threshold of p < .001 (uncorrected), a **positive correlation** of lag-0 cc with activation of the left calcarine gyrus was found during **Subject Extraction** (**Figure 56** and **Table 12**). This means that higher values of lag-0 cc corresponded to increased activation in the left calcarine gyrus.

Activations in the bilateral cerebellum, left fusiform and inferior temporal gyrus were significantly positively correlated with the test scores during the subject extraction phase (**Figure 57** and **Table 13**). No other correlations were significant at the conservative threshold of p < .05 FWE.

Activity in the left supramarginal gyrus was negatively correlated with test scores (**Figure 58** and **Table 14**) during the object encoding.

Sentence Context: The contrast between **checkerboard** repeated and diversified checkerboards [sSC > dSC] did not elicit any significant activations at the stringent threshold of p < .05 FWE. However, at the more liberal threshold of p < .001 significant activations were found in the right inferior frontal gyrus, right angular gyrus and left precuneus. These results are displayed in **Figure 59** and **Table 15**.

Figure 54 - Correlation: checkerboard [lag-0 cc].

Areas displaying a significant negative correlation with the lag-0 cross-correlations are displayed on a render image (p < .05, FWE-corrected for multiple comparisons at the cluster level, k > 30).

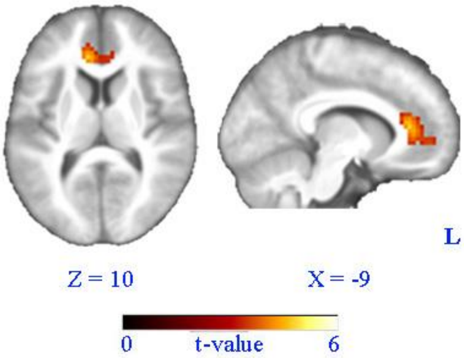


Table 10 - Correlation: checkerboard [lag-0 cc].

Areas displaying a negative correlation with the lag-0 cross-correlations during checkerboard observation (p < 0.05, cluster level, FWE-corrected for multiple comparisons, k > 30). Activations are grouped according to activation clusters; only local maxima are reported.

	Cluster level		Voxel Level	MNI coordinates		
Anatomical Location	**K**	**p-value**	**z-score**	**X**	**Y**	**Z**
L Anterior Cingulate Cortex	167	.020	3.46	-12	47	1

Figure 55 - Correlation: checkerboard [test scores].

Areas displaying a positive correlation with the test-scores are displayed (p < .001, k > 20).

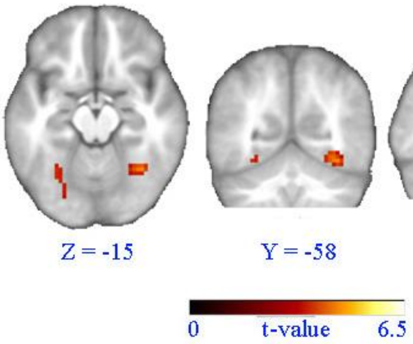


Table 11 - Correlation: checkerboard [test scores].

Areas displaying a positive correlation with the test scores during checkerboard observation (p < 0.001, uncorrected, * = activations surviving an extent threshold criterion, p < .05, k >30). Activations are grouped according to activation clusters; only local maxima are reported.

	Cluster level	Voxel Level		MNI coordinates		
Anatomical Location	**K**	**p-value**	**z-score**	**X**	**Y**	**Z**
LR* Fusiform Gyrus	22	.000	3.38	-27	-64	-11
	39*	.000	3.98	36	-58	-14

Figure 56 - Correlation: subject [lag-0 cc].

Areas displaying a positive correlation with the lag-0 cross-correlations are displayed (p < .001, k > 20).

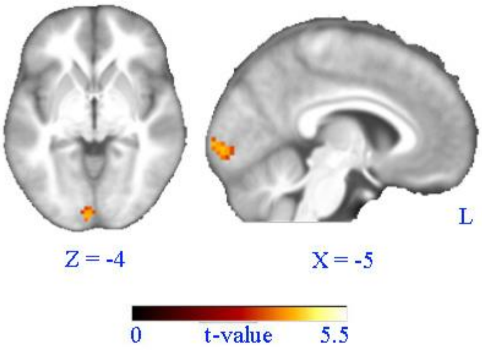


Table 12 - Correlation: subject [lag-0 cc].

Areas displaying a negative correlation with the lag-0 cross-correlations during checkerboard observation (p < 0.001, cluster level). Activations are grouped according to activation clusters; only local maxima are reported.

	Cluster level	Voxel Level		MNI coordinates		
Anatomical Location	**K**	**p-value**	**z-score**	**X**	**Y**	**Z**
L Calcarine Gyrus (BA17)	32	.000	3.79	-6	-94	-2

Figure 57 – Positive correlation subject [test scores].

Areas displaying a positive correlation with the test scores during the subject extraction observation (p < 0.05, cluster level, FWE-corrected for multiple comparisons).

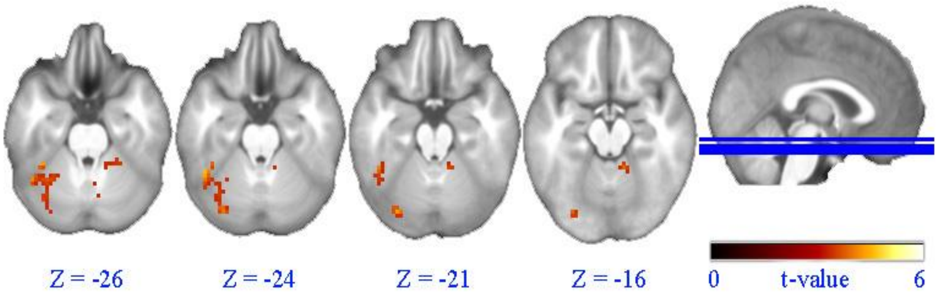


Table 13 - Positive correlation subject [test scores].

Areas displaying a positive correlation with the test scores during the subject extraction observation (p < 0.05, cluster level, FWE-corrected for multiple comparisons). Activations are grouped according to activation clusters; only local maxima are reported.

	Cluster level		Voxel Level	MNI coordinates		
Anatomical Location	**K**	**p-value**	**z-score**	**X**	**Y**	**Z**
Right						
			4.12	24	-40	-32
R Cerebellum	113	.022	4.09	12	-61	-38
			3.52	12	-46	-17
Left						
L Inferior Temporal Gyrus			4.04	-45	-49	-23
L Cerebellum	97	.036	3.86	-30	-79	-20
L Fusiform Gyrus			3.84	-36	-43	-26

Figure 58 - Negative correlation: object [test scores].

Areas displaying a negative correlation with the test scores during the subject extraction observation ($p < 0.05$, $k > 30$).

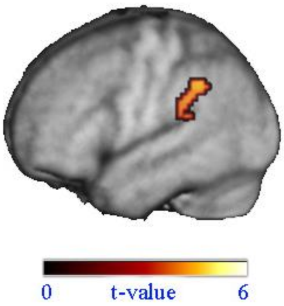


Table 14 -. Negative correlation: object [test scores].

Areas displaying a correlation with the test scores during the object observation ($p < 0.05$, $k > 30$). Activations are grouped according to activation clusters; only local maxima are reported.

	Cluster level	Voxel Level		MNI coordinates		
Anatomical Location	**K**	**p-value**	**z-score**	**X**	**Y**	**Z**
L Supramarginal Gyrus	31	.000	3.49	-63	-37	34
				-66	-34	22

Figure 59 - Contrast: checkerboard [sSC > dSC].

Activations are displayed on axial slices of the average anatomical image of all participants warped to the MNI system of coordinates (p < .001, k > 30).

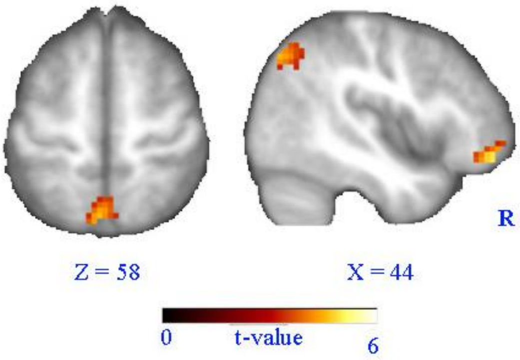


Table 15 – Contrast: checkerboard [sSC>sDC].

Specific activations for the contrast checkerboard [sSC>sDC] (p < 0.001, uncorrected for multiple comparisons, k > 30). Activations are grouped according to activation clusters; only local maxima are reported.

	Cluster level	Voxel Level		MNI coordinates		
Anatomical Location	**K**	**z-score**	**p-value**	**X**	**Y**	**Z**
L Precuneus	97	4.29	.000	-3	-70	55
R Angular Gyrus	58	4.29	.000	42	-70	43
R Inferior Frontal Gyrus (p. Orbitalis)	34	4.35	.000	45	47	-14

3. Non-significant contrasts

Object Encoding, Sentence Context: The contrast **object [sSC > dSC]** did not elicit significant activations at any threshold.

Object Encoding, Social Interaction: There were no suprathreshold clusters for this contrast at any threshold.

Object Encoding, Correlation with Lag-0 cc: No areas displayed a significant correlation with the lag-0 cross-correlations.

Appendix E: Tables

Table 16 - Contrast checkerboard [S+ > S-].

Specific activations for the contrast **checkerboard** [S+ > S-] ($p < 0.001$, uncorrected for multiple comparisons, $k > 10$). Activations are grouped according to activation clusters; only local maxima are reported.

	Cluster level	Voxel Level		MNI coordinates		
Anatomical Location	**K**	**z-score**	**p-value**	**X**	**Y**	**Z**
L Thalamus	15	3.64	.000	-21	-34	4
L Hippocampus				-18	-37	1
R Thalamus (Temporal)	21	3.82	.000	24	-31	1

Table 17 - Correlation: checkerboard [r(lag-0, (S- > S+))].

Areas displaying a positive correlation with the lag-0 cross-correlations during checkerboard observation ($p < 0.001$, cluster level). Activations are grouped according to activation clusters; only local maxima are reported.

	Cluster level	Voxel Level		MNI coordinates		
Anatomical Location	**K**	**p-value**	**z-score**	**X**	**Y**	**Z**
L Middle Temporal Gyrus	24	.000	3.46	-63	-40	7
		.000	3.33	-48	-40	4

Table 18 - Contrast subject [S+ > S-].

Specific activations for the contrast **subject** [S+ > S-] (p < 0.001, uncorrected for multiple comparisons, k > 13). Activations are grouped according to activation clusters; only local maxima are reported.

	Cluster level	Voxel Level		MNI coordinates		
Anatomical Location	**K**	**z-score**	**p-value**	**X**	**Y**	**Z**
L Middle Occipital Gyrus	17	3.98	.000	-36	-64	-5
	28	3.94	.000	-24	-91	-7
L Inferior Temporal Gyrus	13	3.28	.001	-39	-22	-23

Table 19 - Contrast subject [(run3 > run1) * (S+ > S-)].

Specific activations for the contrast **subject[(run3 > run1) * (S+ > S-)]** (p < 0.001, uncorrected for multiple comparisons, k > 13). Activations are grouped according to activation clusters; only local maxima are reported (* = result surviving an extent threshold criterion).

	Cluster level	Voxel Level		MNI coordinates		
Anatomical Location	**K**	**z-score**	**p-value**	**X**	**Y**	**Z**
R Supramarginal Gyrus*	39	3.82	.000	57	-34	37
R Cerebellum	20	3.63	.000	15	-82	32
L Cerebellum	14	3.90	.000	-42	-61	-23

Table 20 - Contrast: checkerboard [dSC > sSC].

Activations are grouped according to activation clusters; only local maxima are reported.

	Cluster level		Voxel Level	MNI coordinates		
Anatomical Location	**K**	**p-value**	**z-score**	**X**	**Y**	**Z**
Right						
R Fusiform Gyrus	213	.000	6.23	27	-49	-11
R Cerebellum			4.79	18	-37	-23
Left						
L Fusiform Gyrus	32	.001	5.05	-27	-52	-11

Table 21 - Sentence Context (p < .05, FWE corrected for multiple comparisons, k>30)

21a) Contrast: subject [sSC > dSC]. Specific activations for the contrast **subject** [sSC > dSC]. Activations are grouped according to activation clusters; only local maxima are reported.

	Cluster level		Voxel Level	MNI coordinates		
Anatomical Location	**K**	**P-value (FWE)**	**z-score**	**X**	**Y**	**Z**
Right						
R Middle Cingulate Cortex	630	.000	7.29	3	-28	43
R Posterior Cingulate Cortex			4.89	12	-46	31
R Superior Temporal Gyrus	2050	.000	7.07	51	-31	22
R Supramarginal Gyrus			6.55	57	-37	31
R Putamen			6.45	30	-4	-2
R Cerebellum (Lobule VI)	169	.000	5.94	27	-49	-29
R Cerebellum (Lobule V)			5.60	15	-40	-23
Left						
L Inferior Parietal Lobule	47	.000	5.49	-51	-61	40
L Supramarginal Gyrus			4.57	-54	-49	34
L Cerebellum (Lobule V)	52	.000	5.27	-24	-52	-26
L Superior Temporal Gyrus	1335	.000	6.30	-63	-28	22
L Amygdala			6.20	-27	-4	-14

L Insula	5.95	-36	-4	13	

21b) Contrast: subject [dSC > sSC]. Specific activations are grouped according to activation clusters; only local maxima are reported.

	Cluster level		Voxel Level	MNI coordinates		
Anatomical Location	**K**	**p-value**	**z-score**	**X**	**Y**	**Z**
Right						
R Inferior Occipital Gyrus	631	.000	6.75	36	-73	11
R Fusiform Gyrus			6.75	30	-64	-8
R Inferior Temporal			6.45	48	-73	-5
R Middle Occipital Gyrus			5.75	39	-76	13
R Calcarine Gyrus			5.10	15	-91	-5
R Inferior Frontal Gyrus (Pars Triangularis)	111	.000	6.10	51	32	19
R Inferior Frontal Gyrus (Pars Opercularis)			5.08	42	14	28
Left						
L Inferior Frontal Gyrus (Pars Triangularis)	464	.000	6.41	-48	23	22
L Inferior Frontal Gyrus (Pars Opercularis)			6.12	-39	8	25
L Inferior Frontal Gyrus (Pars Triangularis			6.10	-48	26	10
L Inferior Frontal Gyrus (Pars Orbitalis)			5.14	-39	32	-14

L Superior Occipital Gyrus			5.69	-9	-94	7
L Fusiform Gyrus	319	.000	5.67	-27	-76	-11
L Lingual Gyrus			5.57	-24	-61	-8

Table 22 - Contrast: object [DC > SC].

Specific activations for the contrast oDC > oSC ($p < 0.05$, cluster level, FWE-corrected for multiple comparisons). Activations are grouped according to activation clusters; only local maxima are reported.

	Cluster level		Voxel Level	MNI coordinates		
Anatomical Location	**K**	**P-value (FWE)**	**z-score**	**X**	**Y**	**Z**
Right						
R Inferior Frontal Gyrus (p. Triangularis)	632	.000	6.11	42	29	19
R Middle Frontal Gyrus			5.80	36	2	40
R Pre-Central Gyrus			5.63	48	2	43
R Thalamus (Parietal)	193	.000	5.52	24	-22	1
R Thalamus (Premotor)			4.70	21	-16	10
Left						
L Inferior Temporal Gyrus	8776	.000	Inf	-48	-55	-11
L Inferior Occipital			7.54	-30	-85	-8
R Middle Occipital			7.49	27	-94	10
L Middle Occipital Gyrus			7.54	-36	-85	4

L Inferior Frontal Gyrus (p. Opercularis)			7.51	-36	8	25
L Inferior Frontal Gyrus (p. triangularis)	1469	.000	6.51	-51	-29	16
L Pre-Central Gyrus			6.90	-45	5	49

Table 23 - Correlation: checkerboard [test scores (dSC (S+ > S-))].

Areas displaying a positive correlation with the test scores during checkerboard observation ($p < 0.05$, corrected). Activations are grouped according to activation clusters; only local maxima are reported.

	Cluster level	Voxel Level		MNI coordinates		
Anatomical Location	**K**	**p-value**	**z-score**	**X**	**Y**	**Z**
R Inferior Frontal Gyrus (p. orb.)	51	.000	3.91	48	35	-11

Table 24 - Correlation: checkerboard [test scores (dSC (S- > S+))].

Areas displaying a positive correlation with the test scores during checkerboard observation ($p < 0.05$, corrected). Activations are grouped according to activation clusters; only local maxima are reported.

	Cluster level	Voxel Level		MNI coordinates		
Anatomical Location	**K**	**p-value**	**z-score**	**X**	**Y**	**Z**
R Lingual Gyrus	74	.000	4.17	12	-70	-11

Table 25 - Correlation: checkerboard [lag0 (sSC (S+ > S-))].

Areas displaying a negative correlation with the lag-0 cross-correlations during checkerboard observation ($p < 0.05$, cluster level, FWE-corrected for multiple comparisons). Activations are grouped according to activation clusters; only local maxima are reported. *denotes activation surviving an extent threshold criterion

	Cluster level		Voxel Level	MNI coordinates		
Anatomical Location	**K**	**p-value**	**z-score**	**X**	**Y**	**Z**
R Caudate Nucleus	126	.041	4.44	12	-1	16
*R Lingual Gyrus	36	.394	3.90	18	-49	4

Table 26 - Contrast subject [(sSC > dSC) * (S+ > S-)].

Specific activations for the contrast **subject [(sSC > dSC) * (S+ > S-)]** (p < 0.001, uncorrected for multiple comparisons, k > 7). Activations are grouped according to activation clusters; only local maxima are reported.

	Cluster level	Voxel Level		MNI coordinates		
Anatomical Location	**K**	**z-score**	**p-value**	**X**	**Y**	**Z**
R Angular Gyrus	16	3.50	.000	48	-64	43
		3.56	.000	48	-61	34
R Middle Frontal Gyrus	7	3.36	.000	33	14	61

Table 27 - Contrast object [(sSC > dSC) * (S+ > S-)].

Specific activations for the contrast **object [(sSC > dSC) * (S+ > S-)]** (p < 0.05, uncorrected for multiple comparisons, k > 30). Activations are grouped according to activation clusters; only local maxima are reported.

	Cluster level	Voxel Level		MNI coordinates		
Anatomical Location	**K**	**z-score**	**p-value**	**X**	**Y**	**Z**
L Cerebellum	39	3.99	.000	-6	-55	-26

Table 28 – Correlation: Independent Components, experimental conditions

Significant results of the one-sample t-test testing for the correlation between each Independent Component and the experimental conditions ($p < 0.05$ FWE).

Table 28a – Component 1: Analysis of the Visual Scene

	Independent Component 1 Analysis of the Visual Scene	
Experimental condition	**P value**	**T value**
cSC	9.3825708e-21	17.964112
cDC	3.6477108e-18	15.148547
sSC	2.2354148e-35	43.687196
sDC	8.9599517e-36	44.717634
oSC	n.s.	--
oDC	0.00016192033	4.1636151

Table 28b – Component 2: complex visuo-spatial encoding.

	Independent Component 2 Complex Visuo-Spatial Encoding	
Experimental condition	**P value**	**T value**
cSC	n.s.	--
cDC	n.s.	--
sSC	1.7717511e-06	5.5915043
sDC	5.1932907e-06	5.2576003
oSC	n.s.	--
oDC	n.s.	--

Table 28c – Component 6: Working memory

	Independent Component 6 Working Memory	
Experimental condition	**P value**	**T value**
cSC	1.0570061e-14	-11.885754
cDC	1.1590782e-13	-10.997851
sSC	n.s.	--
sDC	4.221699e-05	-4.5979705
oSC	n.s.	--
oDC	n.s.	--

Table 28d – Component 7: language network.

	Independent Component 7 Semantic retrieval / Language-network	
Experimental condition	**P value**	**T value**
cSC	1.0045249e-05	5.0515386
cDC	n.s.	_
sSC	1.5001913e-17	14.531317
sDC	3.6546525e-26	25.203965
oSC	5.0253151e-05	-4.5422911
oDC	n.s.	--

Table 28e – Component 8: motor response.

	Independent Component 8 Motor response	
Experimental condition	**P value**	**T value**
cSC	n.s.	--
cDC	n.s.	--
sSC	1.3769578e-18	15.58445
sDC	3.2502909e-19	16.247093
oSC	6.3576189e-09	-7.3387318
oDC	1.2057512e-05	-4.9942891

Table 28f – Component 9: right fronto-parietal attention network.

	Independent Component 9 Right Attention Network	
Experimental condition	**P value**	**T value**
cSC	2.0116423e-07	6.2638782
cDC	0.0012948824	3.4608201
sSC	1.4671336e-08	7.0766682
sDC	6.02075e-06	5.21152
oSC	n.s.	--
oDC	n.s.	--

Table 28g – Component 10

	Independent Component 10 Not identified	
Experimental condition	**P value**	**T value**
cSC	n.s.	--
cDC	0.0025560148	-3.2186102
sSC	4.7500253e-11	8.9153465
sDC	4.4910163e-07	6.0157814
oSC	n.s.	--
oDC	n.s.	--

Table 28h – Component 14: Default Mode Network.

	Independent Component 14 **Default Mode Network**	
Experimental condition	**P value**	**T value**
cSC	4.2651599e-05	-4.5947023
cDC	0.00025512947	-4.0137355
sSC	1.4939917e-13	-10.905897
sDC	3.4957241e-15	-12.30913
oSC	0.012341093	2.6208337
oDC	0.0043093901	3.0268095

Table 28j – Component 20: Social Cognition.

	Independent Component 20 Social Cognition	
Experimental condition	**P value**	**T value**
cSC	1.1471682e-07	6.4375581
cDC	2.2258229e-05	4.8012302
sSC	5.3816471e-18	-14.976998
sDC	1.0816984e-18	-15.693819
oSC	0.021521772	-2.3924853
oDC	1.1259359e-06	-5.7317874

Appendix F: ICA Components

1. Component 1: Visuo-Spatial Analysis

Figure 60 – ICA Component 1.

The first component is overlayed on a mean anatomical image from all participants, warped to the MNI anatomical space.

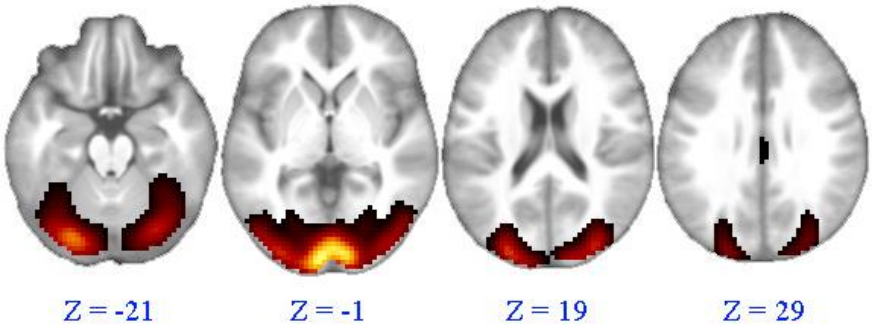


Table 29 – ICA component 1.

Local maxima of the first component are summarized, together with their function in the represented network.

Anatomical Location	BA Areas	Function
Bilateral		
LR Calcarine Gyrus	BA 18 / BA17	Visual Analysis
LR Fusiform Gyrus	HOC4v (V4)	Visual Analysis (secondary)

LR Middle Occipital Gyrus	BA 18	Visual Analysis (secondary)
Right		
R Superior Parietal Lobule	SPL (7a)	Spatial orientation

---

2. Component 3: Complex Visuo-Spatial Encoding

Figure 61 - ICA Component 3.

The third component is overlayed on a mean anatomical image from all participants, warped to the MNI anatomical space.

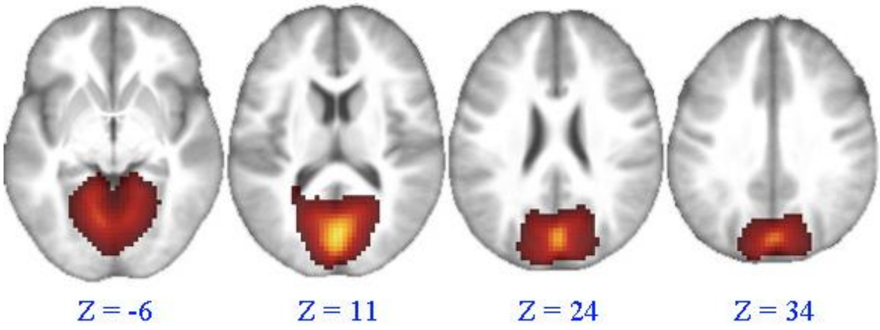


Table 30 – ICA Component 3.

Local maxima of the first component are summarized, together with their function in the represented network.

Anatomical Location	BA Areas	Function
Left		
L Calcarine Gyrus	BA 18	Visual Analysis (Secondary)

L Posterior/Middle Cingulate Cortex	SPL	Spatial Memory Configural Learning
Right		
R Lingual Gyrus	BA 17	Complex Images Encoding Spatial Attention

---

3. Component 6: “Where” - “when” pathway

Figure 62 - ICA Component 6.

The sixth component is overlayed on a mean anatomical image from all participants, warped to the MNI anatomical space.

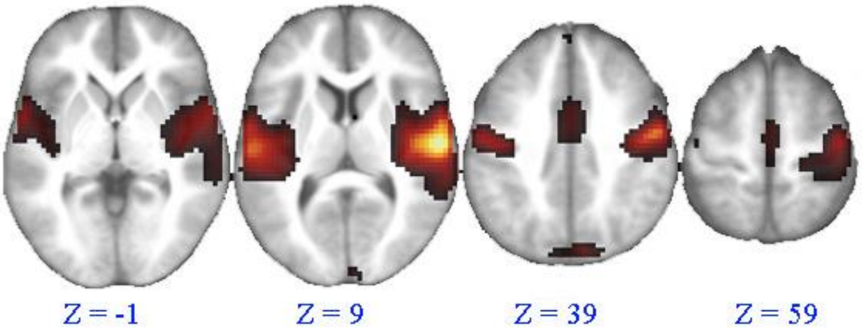


Table 31 – ICA component 6.

Local maxima of the first component are summarized, together with their function in the represented network.

Anatomical Location	BA Areas	Function
Bilateral		
LR Amygdala		Memory, Emotions, Decision Making (limbic system)
LR Post-Central Gyrus	3b 4p	Primary-Secondary sensory
Left		
L Superior Temporal Gyrus	OP1	Laguage, Audition, Social Cognition
L Middle Cingulate gyrus	BA 6	Emotions, Cognitive Control
L Superior Medial Gyrus		Self-awareness
L Inferior Frontal Gyrus (triangularis)	BA 44	Syntax/Semantics
Right		
R Rolandic Operculum	OP1	
R Cuneus		Visual Analysis
R Lingual Gyrus	BA 18	Vision / Word processing
R Paracentral Lobule	SPL (5m)	SMA/S1

4. Component 7: Semantic retrieval / language

Figure 63 - ICA Component 7.

The seventh component is overlayed on a mean anatomical image from all participants,

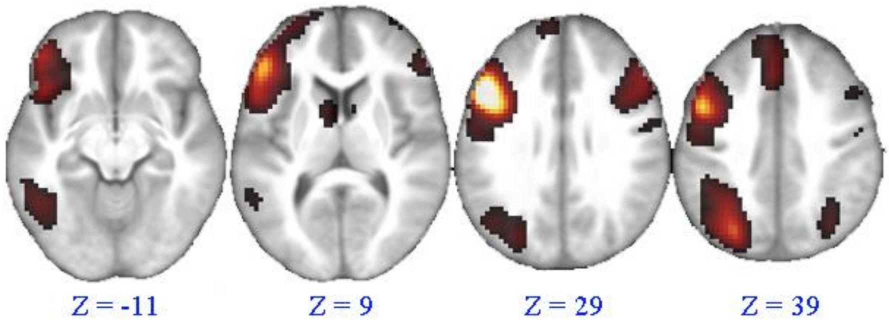


Table 32 - ICA component 7.

Local maxima of the first component are summarized, together with their function in the represented network.

Anatomical Location	BA Areas	Function
Bilateral		
LR Inferior Frontal Gyrus (triangularis)	BA 44 BA 45	Syntax/Semantics
LR Post-Central Gyrus	3b	Sensory
Left		
L SMA	Area 6	Motor Planning
L Middle Frontal Gyrus		
L Inferior Parietal Lobule	IPC	Language, Emotions Senory Analysis
L Inferior Temporal Gyrus		Object Recognition

L Middle Temporal Gyrus		Semantic Memory Words Meaning Recognition of known Faces
Right		
R Angular Gyrus	SPL	Spatial Cognition, Attention, Memory, Social Cognition
R Cerebellum	Lobule VI (Hem)	

5. Component 8: Motor Response

Figure 64 - ICA Component 8.

The eighth component is overlayed on a mean anatomical image from all participants, warped to the MNI anatomical space.

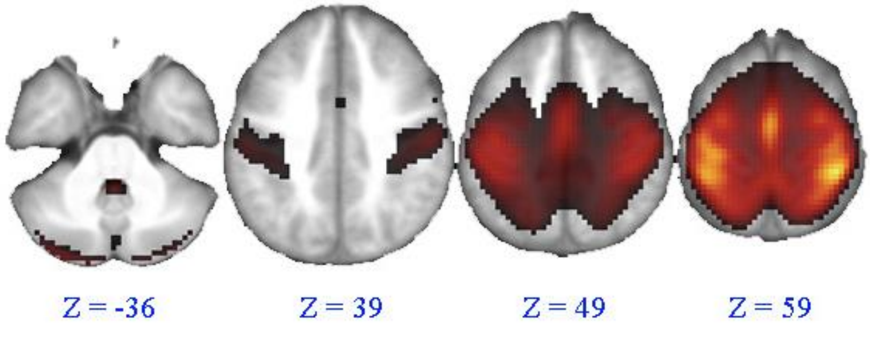


Table 33 - ICA component 8.

Local maxima of the first component are summarized, together with their function in the represented network.

Anatomical Location	BA Areas	Function
Bilateral		
LR Post-Central Gyrus	BA2	Somato – Sensory cortices
LR Cerebellum	VII a	Motor/sensory
LR Pre-Central Gyrus	4a / BA 6	Motor execution
Left		
L Sma	BA 6	Motor Planning

6. Component 9: Right Attentional (Fronto-Parietal)

Figure 65 - ICA Component 9.

The ninth component is overlaid on a mean anatomical image from all participants, warped to the MNI anatomical space.

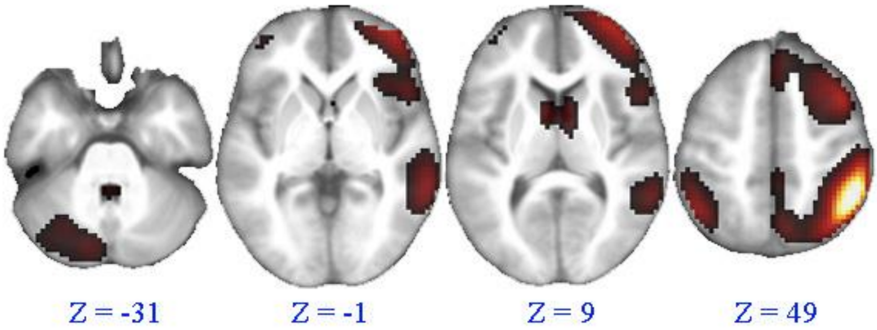


Table 34 - ICA component 9.

Local maxima of the first component are summarized, together with their function in the represented network.

Anatomical Location	BA Areas	Function
Bilateral		
LR Superior Medial gyrus	BA 6	Self-awareness, planning
LR Middle Frontal Gyrus		Self-awareness, planning
LR Inferior Parietal Lobule	IPC (PFm)	Language, Emotions Senory Analysis
Left		
L cerebellum	VIIa	
Right		
R Middle Temporal gyrus		Word processing, Visual processing
R Precuneus	SPL (7a)	Episodic memory, spatial navigation
R Middle Cingulate Cortex		
R inferior frontal gyrus (opercularis)	BA 44	Syntax
R inferior frontal gyrus (orbitalis)		Semantics
R Insula Lobe		Emotions, Social, Awareness

7. Component 10: Working Memory

Figure 66 - ICA Component 10.

The 10th component is overlaid on a mean anatomical image from all participants, warped to the MNI anatomical space.

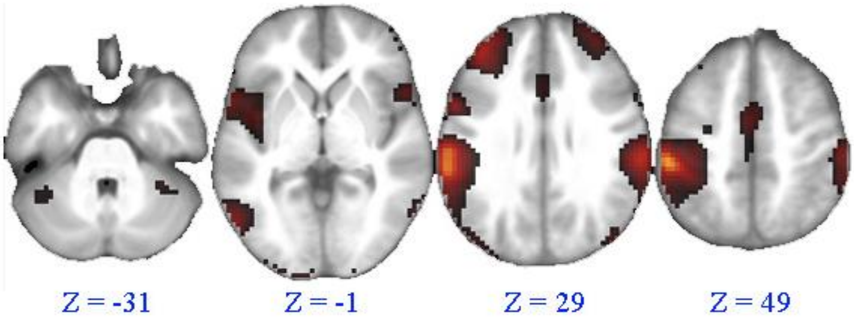


Table 35 - ICA component 10.

Local maxima of the first component are summarized, together with their function in the represented network.

Anatomical Location	BA Areas	Function
Bilateral		
LR Supramarginal gyrus	IPC (PFt)	Empathy, Attention, Language
LR Middle Temporal Gyrus		Word/visual processing
LR Pre-Central Gyrus		Motor execution
LR Middle Frontal Gyrus		Attention, Awareness
LR Cerebellum		
Left		
L Post-Central Gyrus	BA 2	Sensory perception
L Middle Cingulate Cortex	BA 4a	Attention, motor function

L Thalamus		Sensory relay
Right		
R Superior Parietal Lobule		Spatial/Visual processing
R Temporal pole		Social/emotional cognition, memory
R Inferior Frontal gyrus (opercularis)	BA 44	Syntax
R Inferior frontal gyrus (triangularis)		Syntax/Semantics
R Middle Occipital Gyrus	IPC	Visual Processing
R Inferior Occipital Gyrus	BA 17	Visual Perception

---

8. Component 14: Default Mode Network

Figure 67 - ICA Component 14.

The 14th component is overlaid on a mean anatomical image from all participants, warped to the MNI anatomical space.

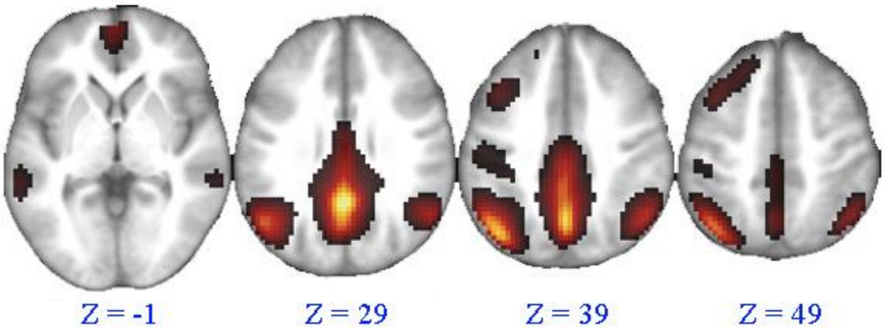


Table 36 – ICA component 14.

Local maxima of the first component are summarized, together with their function in the represented network.

Anatomical Location	BA Areas	Function
Bilateral		
LR precuneus		Episodic Memory, Spatial navigation
LR Mid orbital Gyrus		
LR Angular Gyrus	IPC	Spatial Cognition, Attention, Memory, Social Cognition
LR Middle Temporal Gyrus		Word processing, Visual processing
Left		
L Middle Frontal Gyrus		Attention, Self-awareness
L Inferior Parietal Lobule	BA 2 IPC	Visual/Emotional processing, attention
L Inferior Frontal Gyrus (orbitalis)		Semantics

9. Component 20: Social Cognition

Figure 68 - ICA Component 20.

The 20th component is overlaid on a mean anatomical image from all participants,

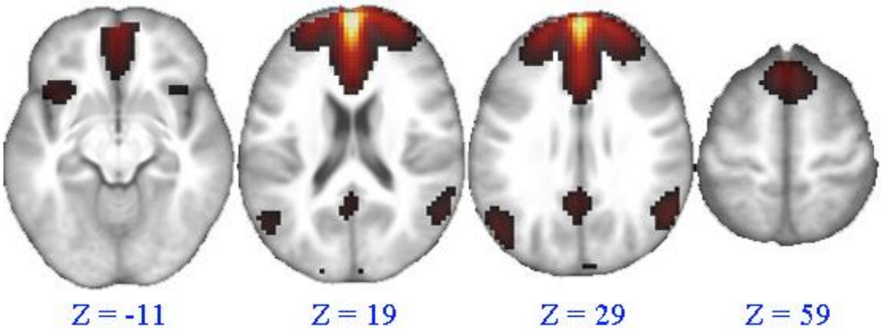


Table 37 – ICA component 20.

Local maxima of the first component are summarized, together with their function in the represented network.

Anatomical Location	BA Areas	Function
Bilateral		
LR Superior Medial Gyrus		Self-awareness
LR Superior Frontal Gyrus		Attention, Self Awareness
LR Middle Cingulate		Attention, Self Awareness
LR Angular Gyrus	IPC	Spatial Cognition, Attention, Memory, Social Cognition
LR Insula		Emotions, Social, Awareness
Left		
L Posterior Cingulate		Emotion, Memory, Default Mode

10. Noise Components (localization of local maxima)

- Component 2: ventricles, bilateral insula
- Component 4: fourth ventricle, third ventricle, lateral ventricle
- Component 5: fourth ventricle
- Component 11: sulci
- Component 12: first maximum in Quadrigeminal Cistern, further maxima in ventricles and sulci.
- Component 13: first maximum in Quadrigeminal Cistern
- Component 15: grey-matter-pia boundary
- Component 16: ventricles, grey-matter-pia boundary
- Component 17: ventricles
- Component 18: ventricles, subcortical noise
- Component 19: grey-matter-pia boundary, ventricles

List of Figures

Chapter 1

Chapter 3

Chapter 4.2

Chapter 4.3

Chapter 4.4

Chapter 4.5

Chapter 5

Appendix D

List Of Tables

List of abbreviations

3T	3 tesla
AC	anterior commissure
ACC	anterior cingulate cortex
AG	angular gyrus
ANCOVA	analysis of covariance
ANOVA	analysis of variance
AoA	age of acquisition
ASD	autistic spectrum disorders
ASL	american sign language
BA	broadmann area
BIA	bilingual interactive activation model
BIA+	latest version of the BIA model
BOLD	blood oxygenetaion level dependent
cc	cross-correlation
CM	concept mediation theory
CV	coefficient of variation
DCM	dynamic causal modelling
DFM	distributed feature model
dpi	dots per inch
dPMC	dorsal Premotor Cortex
dSC	different sentence context
EEG	electroencephalography
EPI	echo-planar imaging
F	female
FDR	false discovery rate
fMRI	functional magnetic resonance imaging
FOV	field of view

FWE	family wise error
GLM	general linear model
IA	interactive activation model
ICA	independent component analysis
IFG	inferior frontal gyrus
iPL	inferior parietal lobule
iPLP	interactive preferential looking paradigm
IPLP	inter-modal preferential looking paradigm
ITI	inter-trial interval
kHz	kilo Hertz
L1	first language
L2	second language
M	mean
M-	non-music
M+	music
MIT	melodic intonation therapy
MNI	montreal neurological institute
MNS	mirror neuron system
mPFC	medial prefrontal cortex
MRI	magnetic resonance imaging
ms	milliseconds
MTG	middle temporal gyrus
MTL	medial temporal lobe
N	sample size
NIRS	near-infrared spectroscopy
PC	posterior commissure
PCA	principal component analysis
pSTS	posterior superior temporal sulcus
pt	points

RF	radiofrequency pulse
RHM	revised hierarchical model
RT	response time
S-	non-social interaction
S+	social interaction
SD	standard deviation
SEM	standard error of the mean
SMA	supplementary motor cortex
SMG	supramarginal gyrus
SPM	statistical parametric map
sSC	same sentence context
T1	spin-lattice relaxation
T2	spin-spin relaxation
TE	echo time
ToM	theory of mind
TP	temporal poles
TPJ	temporo-parietal junction
TR	repetition time
V1	primary visual area
VDM	voxel displacement maps
VOI	volume of interest
vPMC	ventral pre-motor cortex
WA	word association theory
ηp 2	partial eta squared

Curriculum Vitae

Name	Laura Verga
Date of birth	22.02.1985
Place of birth	Busto Arsizio, Italy

Education

11/2014 – present	Research engineer – Centre Euromov, Montpellier (France)
11/2010 - present	PhD student – Max-Planck Institute for Human Cognitive and Brain Science, Leipzig (Germany) & EBRAMUS (Europe BRAin and MUSic Consortium, grant agreement no.238157).
10/2007 - 03/2010	Master of Sciences in Cognitive Neuroscience – Vita-Salute San Raffaele University, Milan (Italy).
10/2004 – 07/2007	Bachelor Degree in Psychology – Vita-Salute San Raffaele University, Milan (Italy).
10/1999 – 07/2004	Science High school "Liceo scientifico A. Tosi", Busto Arsizio (Italy), experimental class in informatics.

Publications

Spada D, **Verga L**, Iadanza A, Tettamanti M, Perani D. (2014) Melody, accompaniment and the auditory scene: an fMRI study with professional pianists. *NeuroImage,* 102(2): 764-775.

Verga, L., & Kotz, S.A. (2013) How relevant is social interaction in second language learning? *Frontiers in Human Neuroscience*, 7:550. doi: 10.3389/fnhum.2013.00550

Publications in preparation

Verga L. & Kotz, S.A. Learning together or learning alone? Social interaction in adult word learning.

Ferreri, L. & **Verga, L**. Music benefits on verbal learning and memory: When and how does it work? A brief review.

Verga, L., Bigand, E. Kotz, S.A. Play along: Temporal coordination to music and social interaction in second language vocabulary learning.

Verga, L. & Kotz, S.A. Neural correlates of second language word learning during social interaction.

English Summary

Learning a new language is a complex and multi-faceted task. As for many other skills, a practical and efficient way of language acquisition often consists in learning from a knowledgeable partner (Frith & Frith, 2012; Hari & Kujala, 2009; Kuhl, 2007). This way, starting with a language building block, namely words, children are able to quickly and effortlessly master a new language in a relatively short period of time (Kuhl et al., 2003). However, learning is a life-long process, which does not end in childhood. Do adults also benefit from the presence of a knowledgeable partner when acquiring a new language? Or are they self-sufficient learners? The aim of the current dissertation is to provide an answer to these questions and to deepen our understanding concerning the role of social interaction in adult word learning.

The introduction of the dissertation presents the theoretical background and empirical evidence motivating the study of word learning during social interaction. First, the rationale for taking interest in the impact of social interaction on cognitive processes is outlined. It is specified how this line of research represents a relatively new focus in cognitive neurosciences (Ochsner & Lieberman, 2001; Schilbach, 2014; Schilbach et al., 2013). Research conducted so far consistently showed that the presence of a human partner may exert a significant influence on how we perceive the world and act in it (Dumas, Nadel, Soussignan, Martinerie, & Garnero, 2010; Hasson, Ghazanfar, Galantucci, Garrod, & Keysers, 2012; Jeong et al., 2010; Knoblich & Sebanz, 2006; Schippers, Gazzola, Goebel, & Keysers, 2009; Vesper, van der Wel, Knoblich, & Sebanz, 2011). Despite this compelling evidence, the complexity of social interactive settings often limit social neuroscience research to simpler cognitive processes leaving open the question as to whether and why social interaction may influence complex cognitive processes such as language learning. Indeed, while the pivotal role of a caregiver in first language learning is now well established (Kuhl, 2003; Tomasello, 2000; Pereira et al., 2008) only a few studies explicitly targeted this matter in adults (Jeong et al., 2010, 2011), while in most cases the importance of social learning is only indirectly hinted at as a venue for future research (Perani et al., 2003; Rodríguez-Fornells et al., 2009).

The second part of the introduction further specifies why social interaction may be of relevance for second language word learning. To this aim, the similarities between learning mechanisms in first and second language acquisition are reviewed. The rationale behind this comparison is that if the processes underlying word acquisition are similar in both infant and adult learners, then it is possible that similar factors – such as social interaction – may exert an influence in both learning situations. In particular, studies and models on first language learning highlight the possibility that a caregiver may help the learner to pinpoint relevant elements in the environment, such as new words' referents (Waxman & Gelman, 2009) by promptly and adaptively coordinating to his behavior (Kuhl, 2007; Pereira et al., 2008). This way, the adult caregiver appears to the child as a particularly salient cue worth giving attention to (Sage & Baldwin, 2010). Accordingly, most models of first language learning emphasize the role of a caregiver in the learning process (Hollich et al., 2000; Kuhl, 2007; Tomasello, 2000). In comparison, models of second language learning focused on variables such as age of acquisition, proficiency, and exposure (Consonni et al., 2013; Perani et al., 2003; Rodríguez-Fornells et al., 2009). While the latter factors have been related to social interaction, the lack of studies directly testing this hypothesis leaves open the possibility that adult learners may not be entirely self-sufficient after all.

The methods part of the dissertation presents and describes the paradigms traditionally employed to investigate word learning and social interaction, respectively, and proposes a novel, unifying method to explore the impact social interaction has on adult word learning. This new method, inspired by game-like paradigms traditionally used in communication research (S. Newman-Norlund et al., 2009; Noordzij et al., 2009; Willems et al., 2010), was developed, validated, and employed in a series of behavioral and neuroimaging (functional magnetic resonance - fMRI) studies. Furthermore, this chapter provides further details about the measures acquired and analyzed in the behavioral studies, followed by an explanation of the basic principles of functional magnetic imaging and of the statistical techniques adopted in the fMRI study (General Linear Model, Independent Component Analysis, Dynamic Causal Modelling).

The empirical part of the dissertation describes five studies. The first two studies were conducted to validate the material that was used in the learning game, consisting of checkerboards containing images depicting simple transitive German sentences, whose object is associated with a novel word. In a first validation step, German native speakers evaluated a set of written German sentences; in the second validation study, a different cohort of native German speakers evaluated the same sentences presented in form of a picture (that is, each element of the sentence - subject, verb, object - was represented by black-and-white pictures). In both experiments, participants rated the sentences' and objects' plausibility; in the second study, they wrote down a sentence conveyed in the pictures. The combined results allowed identifying a set of sentences judged as plausible, whose objects represented good endings for the sentence contexts, and with an optimal agreement between the meaning of the pictures reported by participants and the meaning of the original written sentences.

This set of "pictorial sentences" was implemented in a word learning game, employed in a series of three studies. The first behavioral study aimed at answering the question as of whether social interaction may indeed influence word learning in young healthy adults. To this aim, two groups of German native speakers performed the word learning game: The first group played the game with an experimenter (a social partner), while participants in the second group played the game alone with a computer program. Results show that social interaction influences the behavior of learners; more specifically, participants learning socially tended to temporally coordinate their behavior with a partner. However, this increased coordination per se did not affect immediate recall or recognition of words in the testing phase: Learning was instead influenced by a combined effect of social interaction and variability in the sentence context. It was concluded that temporal coordination may provide a strategy, by which a social partner influences and directs the learner's attention and facilitates the identification of a new word referent, similarly to what has been proposed for first language learning (Kuhl et al., 2003, 2007; Pereira et al., 2008).

The second behavioral study aimed at clarifying whether these results are specific to social interaction, or whether other stimuli with similar properties (e.g., being complex, rhythmic, and potentially interpersonally coordinated – M. Wilson and Wilson, 2005) may also elicit the development of a temporally coordinated behavior in the learner and, as a consequence, facilitate learning. In this study, participants performed the word learning game either alone with a computer, with a social partner, with a music template, or both (that is, a music template *and* a social partner). Results showed that while both music and social interaction influenced the participants' behavior, they had different impact. In particular, when both music and a social partner were presented together, temporal coordination was hindered; this result is suggestive of different mechanisms employed to coordinate to the two stimuli, namely social interaction and music. Whether one or the other approaches is more efficient for word learning, however, is a question still requiring more investigations as no differences were observed between conditions in a retrieval phase taking place immediately after the learning session. While contributing to the literature on word learning in adults by investigating two possible factors influencing adult word learning, this study also has important implications for situations – such as music therapy - in which music and social interaction are present at the same time.

The final study investigated the neural basis of adult word learning during social interaction (or a “second person approach”, Schilbach et al., 2013; Schilbach, 2014). Previous studies have shown that the interaction with a social partner boosts brain activations in those brain areas that are recruited for the task at hand, such as areas involved in word learning tasks (Jeong et al., 2010). Given the two previous studies reported here, the possibility emerges that a social partner, as a responsive agent, may impact the learner’s attention. Results of the last study of this dissertation show that social interaction significantly modulates the activation in areas involved not only in word learning, but also in visuo-spatial attention. Moreover, the connectivity of attention-related areas is increased by social interaction. However, this modulation depends upon task specifics: If the task is easy enough for the learner (for example, if the sentence context has a low degree of variability), no differences are observed between the social and the non-social group. These results suggest that social interaction may influence visuo-spatial attention during adult second language word learning in combination with the characteristics of the sentence context.

Take together, the results of the behavioral studies provide first evidence that social interaction influences adults while learning new words of a second language. More specifically, adults tend to temporally coordinate their behavior with a partner; this effect may represent a strategy by which a knowledgeable partner directs the learner’s attention towards relevant elements in the environments (that is, new words referents). Furthermore, this effect may be specific to social interaction, as other stimuli with similar properties (music) do not influence participants in the same way. The fMRI study further supports and extends this claim, by showing how social interaction impacts activation in task-related areas (areas related to word learning and visuo-spatial attention) and their connectivity. In addition, the three studies consistently report that social interaction acts in concert with the characteristics of the sentence context a new word is embedded in: While social interaction is certainly used by adult learners, contextual elements also play a pivotal role.

Based on the data collected in the studies of this dissertation, a model of social word learning in healthy adults is proposed in the last part of the dissertation; this model suggests that word learning may follow two different routes: A direct route and an indirect route. The direct route is used when the task is easy enough for the participant to learn in a game context by her/himself. In this scenario, the learner is able to identify a referent and to map its meaning onto the new word directly, without requiring external help. However, when the learning environment is more complex, the presence of another person simplifies the identification of a word's referent, hence facilitating learning. This indirect route benefits from strategies reported in the empirical studies, such as the temporal coordination to direct the learner's attention.

In summary, the current dissertation provides first evidence of the role played by social interaction in second language adult word learning. More specifically, it suggests that adults may be defined as utilitarian social learners, who benefit from the presence of a social partner as long as a task is too challenging to be performed alone. These results motivate a novel model of social word learning, that allows generating testable predictions for future studies in both healthy and clinical populations (for example, aphasic patients that need to re-learn words).

Zusammenfassung

Eine neue Sprache zu erlernen ist eine komplexe und vielschichtige Aufgabe. Wie bei vielen anderen Fertigkeiten, besteht ein praktischer und effizienter Weg für den Spracherwerb im Lernen von einem sachkundigen Partner (Frith & Frith, 2012; Hari & Kujala, 2009; Kuhl, 2007). Auf diese Art, beginnend mit Sprachbausteinen, nämlich Wörtern, sind Kinder in der Lage, eine neue Sprache in relativ kurzer Zeit schnell und ohne Aufwand zu meistern (Kuhl et al., 2003). Lernen ist jedoch ein lebenslanger Prozess, der nicht in der Kindheit endet. Profitieren auch Erwachsene von der Anwesenheit eines sachkundigen Partners, wenn sie eine neue Sprache erwerben? Oder sind sie autarke Lernende? Ziel dieser Dissertation ist, diese Fragen zu beantworten und unser Verständnis bezüglich der Rolle sozialer Interaktion im Erlernen von Wörtern bei Erwachsenen zu vertiefen.

Die Einleitung der Dissertation erläutert den theoretischen Hintergrund und empirische Evidenz, die die Untersuchung des Wortlernens während sozialer Interaktion motivieren. Zunächst werden die Gründe für das Interesse am Einfluss sozialer Interaktion auf kognitive Prozesse skizziert. Es wird spezifiziert, inwiefern diese Forschung einen relativ neuen Fokus innerhalb der kognitiven Neurowissenschaften darstellt (Ochsner & Lieberman, 2001; Schilbach, 2014; Schilbach et al., 2013). Bisherige Forschung hat in konsistenter Weise gezeigt, dass die Anwesenheit eines menschlichen Partners einen signifikanten Einfluss auf unser Verhalten in und unsere Wahrnehmung der Welt ausüben kann (Dumas, Nadel, Soussignan, Martinerie, & Garnero, 2010; Hasson, Ghazanfar, Galantucci, Garrod, & Keysers, 2012; Jeong et al., 2010; Knoblich & Sebanz, 2006; Schippers, Gazzola, Goebel, & Keysers, 2009; Vesper, van der Wel, Knoblich, & Sebanz, 2011). Trotz dieser triftigen Evidenz beschränkt die Komplexität sozialerinteraktiver Settings sozial-neurowissenschaftliche Forschung oftmals auf simple kognitive Prozesse, wobei die Frage offenbleibt, ob und warum soziale Interaktion komplexe kognitive Prozesse wie Sprachenlernen beeinflusst. Während die Schlüsselrolle einer Bezugsperson beim Erstspracherwerb mittlerweile weithin anerkannt wird (Kuhl, 2003; Tomasello, 2000; Pereira et al., 2008), haben nur wenige Studien diesen Sachverhalt bei Erwachsenen untersucht (Jeong et al., 2010, 2011), wobei in den meisten Fällen auf die Wichtigkeit sozialen Lernens nur indirekt als zukünftiges Forschungsthema hingewiesen wird (Perani et al., 2003; Rodríguez-Fornells et al., 2009).

Der zweite Teil der Einleitung spezifiziert weitergehend, warum soziale Interaktion für das Erlernen von Wörtern beim Zweitspracherwerb relevant sein könnte. Zu diesem Zweck werden die Ähnlichkeiten zwischen Lernmechanismen für den Erst- und Zweitspracherwerb rezensiert. Der Grund für diesen Vergleich ist, dass, wenn die Prozesse des Worterwerbs zwischen kindlichen und erwachsenen Lernenden ähnlich sind, es möglich ist, dass ähnliche Faktoren – wie soziale Interaktion – einen Einfluss auf beide Lernsituationen ausüben. Insbesondere heben Studien und Modelle des Erstspracherwerbs die Möglichkeit hervor, dass eine Bezugsperson dem Lernenden helfen könnte, relevante Elemente in der Umwelt zu lokalisieren, beispielsweise Referenten neuer Worte (Waxman & Gelman, 2009), indem sie sich unmittelbar und adaptiv auf sein Verhalten ausrichten (Koordination) (Kuhl, 2007; Pereira et al., 2008). Auf diese Weise erscheint die erwachsene Bezugsperson als besonders salienter Hinweisreiz, der es Wert ist, dass man ihm seine Aufmerksamkeit zuwendet (Sage & Baldwin, 2010). Dementsprechend betonen die meisten Modelle erstsprachlichen Lernens die Rolle der Bezugsperson im Lernprozess (Hollich et al., 2000; Kuhl, 2007; Tomasello, 2000). Im Vergleich dazu fokussierten Modelle zweitsprachlichen Lernens auf Variablen wie Alter des Erwerbs, Sprachbeherrschung und in welchem Umfang der Lernende der Sprache ausgesetzt ist (Exposition) (Consonni et al., 2013; Perani et al., 2003; Rodríguez-Fornells et al., 2009).

Während letztere Faktoren zu sozialer Interaktion in Bezug gesetzt wurden, lässt das Fehlen von Studien, die diese Hypothese direkt testen, die Möglichkeit offen, dass erwachsene Lernende letztlich nicht vollständig autark sind.

Der Methodenteil der Dissertation beschreibt die traditionell angewandten Paradigmen zur Untersuchung des Erlernens von Wörtern bzw. sozialer Interaktion, und schlägt eine neue, vereinende Methode zur Untersuchung des Einflusses sozialer Interaktion auf adultes Wortlernen vor. Diese neue Methode, inspiriert durch in der Kommunikationsforschung traditionell benutzte spielartige Paradigmen (Newman-Norlund et al., 2009; Noordzij et al., 2009; Willems et al., 2010), wurde entwickelt, validiert und angewandt in einer Reihe behavioraler und bildgebender (fMRT-)Studien. Weiterhin werden in diesem Kapitel die in den behavioralen Studien gewonnenen und analysierten Messungen detaillierter ausgeführt, gefolgt von einer Erklärung der grundlegenden Prinzipien funktionaler magnetischer Bildgebung und der in der fMRT-Studie verwendeten statistischen Techniken (General Linear Model, Independent Component Analysis, Dynamic Causal Modelling).

Der empirische Teil der Dissertation erläutert fünf Studien. Die ersten beiden Studien wurden durchgeführt, um das im Lernspiel verwendete Material zu validieren, das aus Schachbrettmustern mit Bildern bestand, die einfache transitive deutsche Sätze darstellten, deren Objekt mit einem neuen Wort assoziiert ist. In einem ersten Schritt der Validierung evaluierten deutsche Muttersprachler ein Set geschriebener deutscher Sätze; in der zweiten Validierungsstudie evaluierte eine andere Kohorte deutscher Muttersprachler dieselben Sätze, die in Form von Bildern dargestellt waren (d.h., jedes Element des entsprechenden Satzes – Subjekt, Verb, Objekt – wurde durch Schwarz-Weiß-Bilder repräsentiert). In beiden Experimenten führten die Probanden Ratings bezüglich der Plausibilität der Sätze und Objekte durch; in der zweiten Studie schrieben sie einen über das Bild vermittelten Satz nieder. Die kombinierten Resultate erlaubten, ein Set von Sätzen zu identifizieren, die als plausibel beurteilt wurden, deren Objekte gute Abschlüsse des Satzkontextes darstellten und die optimale Übereinstimmung zwischen der von den Probanden berichteten Bedeutung der Bilder und der tatsächlichen Bedeutung der geschriebenen Sätze gewährleisteten.

Dieses Set „bildhafter Sätze“ wurde in ein Wortlernspiel implementiert, das in einer Reihe dreier Studien verwendet wurde. Die erste Verhaltensstudie zielte darauf ab, die Frage zu beantworten, ob soziale Interaktion tatsächlich das Erlernen von Wörtern bei jungen gesunden Probanden beeinflussen könnte. Zu diesem Zweck führten zwei Gruppen deutscher Muttersprachler das Wortlernspiel aus: die erste Gruppe spielte das Spiel mit einem Experimentator (einem Sozialpartner), während die Probanden der zweiten Gruppe das Spiel allein mit einem Computerprogramm spielten. Die Resultate zeigen, dass soziale Interaktion das Verhalten der Lernenden beeinflusst; genauer gesagt tendierten sozial lernende Probanden dazu, ihr Verhalten zeitlich mit einem Partner zu koordinieren. Diese gesteigerte Koordination beeinflusste jedoch nicht per se den sofortigen Abruf oder die Erinnerung an die Wörter in der Testphase, das Lernen war stattdessen von einem kombinierten Effekt von sozialer Interaktion und Variabilität im Satzkontext beeinflusst. Es wurde gefolgert, dass zeitliche Koordination eine Strategie zur Verfügung stellen könnte, durch die ein Sozialpartner die Aufmerksamkeit des Lernenden beeinflusst und lenkt und die Identifikation des Referenten eines neuen Wortes erleichtert, was in ähnlicher Weise für den Erstsprachenerwerb vorgeschlagen wurde (Kuhl et al., 2003, 2007; Pereira et al., 2008).

Die zweite Verhaltensstudie hatte zum Ziel, zu klären, ob diese Resultate spezifisch für soziale Interaktion waren, oder ob andere Stimuli mit ähnlichen Eigenschaften (z.B. komplex, rhythmisch und potentiell interpersonell koordiniert zu sein – M. Wilson & Wilson, 2005) auch die Entwicklung zeitlich koordinierten Verhaltens im Lernenden hervorrufen könnte und, als Konsequenz, Lernen erleichtert. In dieser Studie führten die Probanden das

Wortlernspiel entweder allein mit einem Computer, mit einem Sozialpartner, einer musikalische Vorlage oder beidem durch (also einer musikalischen Vorlage und einem Sozialpartner). Die Resultate zeigten, dass Musik und soziale Interaktion, obwohl beide das Verhalten des Probanden beeinflussten, unterschiedliche Auswirkungen hatten. Insbesondere war die zeitliche Koordination behindert, wenn Musik und soziale Interaktion zusammen präsentiert wurden. Dieses Ergebnis suggeriert unterschiedliche Mechanismen, die verwendet werden, um die zwei Stimuli, nämlich soziale Interaktion und Musik, zu koordinieren. Ob der eine oder der andere Zugang für das Wortlernen effizienter ist, bleibt eine Frage, die mehr Untersuchungen erfordert, da keine Unterschiede zwischen den Bedingungen in der Phase des Wortabrufs, die sofort nach der Lernphase stattfand, beobachtet wurde. Neben der Tatsache, dass diese Studie zur Literatur über Wortlernen bei Erwachsenen durch die Untersuchung zweier möglicher Einflussfaktoren auf adultes Wortlernen beiträgt, weist sie auch bedeutsame Implikationen für Situationen auf, in der Musik und soziale Interaktion zur selben Zeit präsent sind, beispielsweise in der Musiktherapie.

Die letzte Studie untersuchte die neuronale Basis adulten Wortlernens während sozialer Interaktion (bzw. eines „zweitpersonalen Zugangs“, Schilbach et al., 2013; Schilbach, 2014). Vorhergehende Studien zeigten, dass die Interaktionen mit einem Sozialpartner Hirnaktivierungen in Arealen verstärkt, die für die augenblickliche Aufgabe benötigt werden, wie Areale, die in Wortlernaufgaben involviert sind (Jeong et al., 2010). Gegeben die beiden hier zuvor vorgestellten Studien, ergibt sich die Möglichkeit, dass ein Sozialpartner - als responsiver Akteur - die Aufmerksamkeit des Lernenden beeinflusst. Die Ergebnisse der letztens Studie dieser Dissertation zeigen, dass soziale Interaktion auf signifikante Weise die Aktivierung in Arealen moduliert, die nicht nur am Erlernen von Wörtern sondern auch an visuell-räumlicher Aufmerksamkeit beteiligt sind. Weiterhin wird die Konnektivität aufmerksamkeitsbezogener Areale durch soziale Interaktion erhöht. Diese Modulation ist jedoch abhängig von Spezifiken der Aufgabe: Ist die Aufgabe für den Lernenden genügend einfach (beispielsweise, wenn der Satzkontext einen geringen Variabilitätsgrad aufweist), sind keine Unterschiede zwischen sozialer und nicht-sozialer Gruppe zu beobachten. Diese Ergebnisse deuten darauf hin, dass soziale Interaktion visuell-räumliche Aufmerksamkeit während adultem zweitsprachlichen Wortlernen in Kombination mit den Charakteristika des Satzkontext beeinflussen könnte.

Zusammengefasst stellen die Ergebnisse der Verhaltensstudien erste Evidenz dar, dass soziale Interaktion Erwachsene beeinflusst, wenn diese neue Worte einer Zweitsprache lernen. Genauer gesagt tendieren Erwachsene dazu, ihr Verhalten zeitlich mit einem Partner zu koordinieren; dieser Effekt repräsentiert möglicherweise eine Strategie, durch die ein sachkundiger Partner die Aufmerksamkeit des Lernenden auf ein relevantes Element in der Umwelt lenkt (d.h. Referenten neuer Wörter). Des Weiteren könnte dieser Effekt spezifisch für soziale Interaktion sein, da andere Stimuli mit ähnlichen Eigenschaften (Musik) Probanden nicht auf die selbe Art beeinflussen. Die fMRT-Studie stützt und erweitert diese Behauptung, indem sie zeigt, wie soziale Interaktion sich auf die Aktivierung aufgabenbezogener Areale (Areale mit Bezug zum Erlernen von Wörtern und zu visuell-räumlicher Aufmerksamkeit) und deren Konnektivität auswirkt. Zusätzlich zeigen die drei Studien konsistent, dass soziale Interaktion gemeinsam agiert mit den Charakteristiken eines Satzkontexts, in den ein neues Wort eingefügt ist: Während soziale Interaktion mit Sicherheit von adulten Lernenden genutzt wird, spielen kontextuelle Elemente ebenfalls eine zentrale Rolle.

Basierend auf den in den Studien dieser Dissertation erhobenen Daten wird im letzten Teil der Dissertation ein Modell sozialen Wortlernens bei gesunden Erwachsenen vorgeschlagen, welches suggeriert, dass das Erlernen von Wörtern zwei verschiedenen Routen folgt: einer direkten Route und einer indirekten Route. Die direkte Route wird genutzt, wenn die Aufgabe für den Probanden genügend einfach ist, um im Kontext eines Spiels selbstständig zu lernen. In diesem Szenario ist der Lernende in der Lage, einen Referenten zu identifizieren und dessen Bedeutung direkt auf ein neues Wort zu übertragen, ohne externe Hilfe zu benötigen. Ist jedoch die Lernumgebung komplexer, vereinfacht die Anwesenheit einer anderen Person die Identifikation des Referenten eines Wortes und erleichtert so das Lernen. Diese indirekte Route profitiert von Strategien, die in den empirischen Studien dargestellt werden, wie zeitlicher Koordination, um die Aufmerksamkeit des Lernenden zu lenken.

Zusammengefasst liefert diese Dissertation erste Evidenz für die Rolle sozialer Interaktion im zweitsprachlichen Wortlernen bei Erwachsenen. Insbesondere deutet sie darauf hin, dass Erwachsene als utilitaristische soziale Lernende definiert werden können, die von der Anwesenheit eines Sozialpartners profitieren, solange die Aufgabe zu schwierig ist, um allein ausgeführt werden zu können. Diese Resultate motivieren ein neuartiges Modell des sozialen Wortlernens sowohl in gesunden, als auch klinischen Populationen (beispielsweise aphasische Patienten, die Wörter neu erlernen müssen).

Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche in dieser Arbeit gekennzeichnet. Die Arbeit ist in keinem früheren Promotionsverfahren angenommen oder abgelehnt worden.

Laura Verga

Leipzig, 06.02.2015

Bibliographic details

Laura Verga

Learning together or learning alone: Investigating the role of social interaction in second language word learning

Fakultät für Biowissenschaften, Pharmazie und Psychologie der Universität Leipzig
Dissertation

358 pages, 439 references, 68 figures, 37 tables

Paper. Do adults benefit from the presence of a social partner when learning a new language? Only a few studies investigated this question. Hence, aim of this dissertation was to deepen our knowledge of the mechanisms supporting word learning in adults during social interaction. In chapter 1 the theoretical background, the empirical issues and the open questions underlying the study of social word learning in first and second language are outlined. chapter 2 describes the research questions addressed in this thesis. chapter 3 describes a method created ex-novo and employed in our experiments. Study 1 and 2, exposed in chapter 4, illustrate the process of validation of this novel paradigm. Study 3 uses this set-up to answer the question as of whether social interaction influences second language learning. Study 4 aims at clarifying whether the effects of a social partner on word learning are unique, or whether other stimuli may exert a similar influence on adult learners. Lastly, study 5 investigates the neural basis of word learning in a social as compared to a non-social setting. Chapter 6 summarizes and discusses the results of all studies, showing how social interaction a) influences the learner's behavior, b) has a unique attentional effect, different from other stimuli, c) influences the activity and connectivity of areas involved in word learning and attention and d) interacts with the characteristics of the context of a word presentation. Based on these results, a new model of social word learning is proposed, together with possible venues for future research and clinical applications.

Referat. Profitieren Erwachsene von der Anwesenheit eines sozialen Partners, wenn sie eine neue Sprache erlernen? Nur wenige Studien haben diese Frage untersucht. Folglich war das Ziel dieser Dissertation, unser Wissen über die Mechanismen zu vertiefen, die das Lernen von Wörtern bei Erwachsenen während sozialer Interaktionen unterstützen. In Kapitel 1 werden der theoretische Hintergund, empirische Belange und offene Fragen umrissen, die der Untersuchung sozialen Wortlernens in Erst- und Zweitsprache zugrunde liegen. Kapitel 2 beschreibt die in vorliegender Arbeit behandelten Forschungsfragen. Kapitel 3 beschreibt eine ex-novo entwickelte und in unseren Experimenten angewandte Methode. Studien 1 und 2, ausgeführt in Kapitel 4, illustrieren den Prozess der Validierung dieses neuen Paradigmas. Studie 3 nutzt dieses Set-up, um die Frage zu beantworten, ob soziale Interaktion das Erlernen einer Zweitsprache beeinflusst. Studie 4 zielt darauf ab, zu klären, ob die Effekte eines Sozialpartners auf das Wortlernen singulär sind, oder ob andere Stimuli möglicherweise einen ähnlichen Einfluss auf erwachsene Lernende ausüben. Schließlich untersucht Studie 5 die neuronale Basis des Wortlernens in sozialen gegenüber nichtsozialen Situationen. Kapitel 6 fasst die Ergebnisse aller Studien zusammen, diskutiert diese und zeigt auf, wie soziale Interaktion a) das Verhalten des Lernenden beeinflusst, b) einen singulären attentionalen Effekt ausübt, der sich von anderen Stimuli unterscheiden lässt, c) die an Wortlernen und Aufmerksamkeit beteiligten Areale hinsichtlich ihrer Aktivität und Konnektivität beeinflusst und d) mit den kontextuellen Charakteristika der Präsentation von Wörtern interagiert. Basierend auf diesen Ergebnissen werden ein neues Modell sozialen Wortlernens sowie mögliche Implikationen für zukünftige Forschung und klinische Anwendungen vorgeschlagen.

MPI Series in Human Cognitive and Brain Sciences:

1 Anja Hahne
Charakteristika syntaktischer und semantischer Prozesse bei der auditiven Sprachverarbeitung: Evidenz aus ereigniskorrelierten Potentialstudien

2 Ricarda Schubotz
Erinnern kurzer Zeitdauern: Behaviorale und neurophysiologische Korrelate einer Arbeitsgedächtnisfunktion

3 Volker Bosch
Das Halten von Information im Arbeitsgedächtnis: Dissoziationen langsamer corticaler Potentiale

4 Jorge Jovicich
An investigation of the use of Gradient- and Spin-Echo (GRASE) imaging for functional MRI of the human brain

5 Rosemary C. Dymond
Spatial Specificity and Temporal Accuracy in Functional Magnetic Resonance Investigations

6 Stefan Zysset
Eine experimentalpsychologische Studie zu Gedächtnisabrufprozessen unter Verwendung der funktionellen Magnetresonanztomographie

7 Ulrich Hartmann
Ein mechanisches Finite-Elemente-Modell des menschlichen Kopfes

8 Bertram Opitz
Funktionelle Neuroanatomie der Verarbeitung einfacher und komplexer akustischer Reize: Integration haemodynamischer und elektrophysiologischer Maße

9 Gisela Müller-Plath
Formale Modellierung visueller Suchstrategien mit Anwendungen bei der Lokalisation von Hirnfunktionen und in der Diagnostik von Aufmerksamkeitsstörungen

10 Thomas Jacobsen
Characteristics of processing morphological structural and inherent case in language comprehension

11 Stefan Kölsch
Brain and Music
A contribution to the investigation of central auditory processing with a new electrophysiological approach

12 Stefan Frisch
Verb-Argument-Struktur, Kasus und thematische Interpretation beim Sprachverstehen

13 Markus Ullsperger
The role of retrieval inhibition in directed forgetting – an event-related brain potential analysis

14 Martin Koch
Measurement of the Self-Diffusion Tensor of Water in the Human Brain

15 Axel Hutt
Methoden zur Untersuchung der Dynamik raumzeitlicher Signale

16 Frithjof Kruggel
Detektion und Quantifizierung von Hirnaktivität mit der funktionellen Magnetresonanztomographie

17 Anja Dove
Lokalisierung an internen Kontrollprozessen beteiligter Hirngebiete mithilfe des Aufgabenwechselparadigmas und der ereigniskorrelierten funktionellen Magnetresonanztomographie

18 Karsten Steinhauer
Hirnphysiologische Korrelate prosodischer Satzverarbeitung bei gesprochener und geschriebener Sprache

19 Silke Urban
Verbinformationen im Satzverstehen

20 Katja Werheid
Implizites Sequenzlernen bei Morbus Parkinson

21 Doreen Nessler
Is it Memory or Illusion? Electrophysiological Characteristics of True and False Recognition

22 Christoph Herrmann
Die Bedeutung von 40-Hz-Oszillationen für kognitive Prozesse

23 Christian Fiebach
Working Memory and Syntax during Sentence Processing.
A neurocognitive investigation with event-related brain potentials and functional magnetic resonance imaging

24 Grit Hein
Lokalisation von Doppelaufgabendefiziten bei gesunden älteren Personen und neurologischen Patienten

25 Monica de Filippis
Die visuelle Verarbeitung unbeachteter Wörter. Ein elektrophysiologischer Ansatz

26 Ulrich Müller
Die katecholaminerge Modulation präfrontaler kognitiver Funktionen beim Menschen

27 Kristina Uhl
Kontrollfunktion des Arbeitsgedächtnisses über interferierende Information

28 Ina Bornkessel
The Argument Dependency Model: A Neurocognitive Approach to Incremental Interpretation

29 Sonja Lattner
Neurophysiologische Untersuchungen zur auditorischen Verarbeitung von Stimminformationen

30 Christin Grünewald
Die Rolle motorischer Schemata bei der Objektrepräsentation: Untersuchungen mit funktioneller Magnetresonanztomographie

31 Annett Schirmer
Emotional Speech Perception: Electrophysiological Insights into the Processing of Emotional Prosody and Word Valence in Men and Women

32 André J. Szameitat
Die Funktionalität des lateral-präfrontalen Cortex für die Verarbeitung von Doppelaufgaben

33 Susanne Wagner
Verbales Arbeitsgedächtnis und die Verarbeitung ambiger Wörter in Wort- und Satzkontexten

34 Sophie Manthey
Hirn und Handlung: Untersuchung der Handlungsrepräsentation im ventralen prämotorischen Cortex mit Hilfe der funktionellen Magnet-Resonanz-Tomographie

35 Stefan Heim
Towards a Common Neural Network Model of Language Production and Comprehension: fMRI Evidence for the Processing of Phonological and Syntactic Information in Single Words

36 Claudia Friedrich
Prosody and spoken word recognition: Behavioral and ERP correlates

37 Ulrike Lex
Sprachlateralisierung bei Rechts- und Linkshändern mit funktioneller Magnetresonanztomographie

38 Thomas Arnold
Computergestützte Befundung klinischer Elektroenzephalogramme

39 Carsten H. Wolters
Influence of Tissue Conductivity Inhomogeneity and Anisotropy on EEG/MEG based Source Localization in the Human Brain

40 Ansgar Hantsch
Fisch oder Karpfen? Lexikale Aktivierung von Benennungsalternative bei der Objektbenennung

41 Peggy Bungert
Zentralnervöse Verarbeitung akustischer Informationen
Signalidentifikation, Signallateralisation und zeitgebundene Informationsverarbeitung bei Patienten mit erworbenen Hirnschädigungen

42 Daniel Senkowski
Neuronal correlates of selective attention: An investigation of electrophysiological brain responses in the EEG and MEG

43 Gert Wollny
Analysis of Changes in Temporal Series of Medical Images

S 1 Markus Ullsperger & Michael Falkenstein
Errors, Conflicts, and the Brain Current Opinions on Performance Monitoring

44 Angelika Wolf
Sprachverstehen mit Cochlea-Implantat: EKP-Studien mit postlingual ertaubten erwachsenen CI-Trägern

45 Kirsten G. Volz
Brain correlates of uncertain decisions: Types and degrees of uncertainty

46 Hagen Huttner
Magnetresonanztomographische Untersuchungen über die anatomische Variabilität des Frontallappens des menschlichen Großhirns

47 Dirk Köster
Morphology and Spoken Word Comprehension: Electrophysiological Investigations of Internal Compound Structure

48 Claudia A. Hruska
Einflüsse kontextueller und prosodischer Informationen in der auditorischen Satzverarbeitung: Untersuchungen mit ereigniskorrelierten Hirnpotentialen

49 Hannes Ruge
Eine Analyse des raum-zeitlichen Musters neuronaler Aktivierung im Aufgabenwechselparadigma zur Untersuchung handlungssteuernder Prozesse

50 Ricarda I. Schubotz
Human premotor cortex: Beyond motor performance

51 Clemens von Zerssen
Bewusstes Erinnern und falsches Wiedererkennen: Eine funktionelle MRT Studie neuroanatomischer Gedächtniskorrelate

52 Christiane Weber
Rhythm is gonna get you.
Electrophysiological markers of rhythmic processing in infants with and without risk for Specific Language Impairment (SLI)

53 Marc Schönwiesner
Functional Mapping of Basic Acoustic Parameters in the Human Central Auditory System

54 Katja Fiehler
Temporospatial characteristics of error correction

55 Britta Stolterfoht
Processing Word Order Variations and Ellipses: The Interplay of Syntax and Information Structure during Sentence Comprehension

56 Claudia Danielmeier
Neuronale Grundlagen der Interferenz zwischen Handlung und visueller Wahrnehmung

57 Margret Hund-Georgiadis
Die Organisation von Sprache und ihre Reorganisation bei ausgewählten, neurologischen Erkrankungen gemessen mit funktioneller Magnetresonanztomographie – Einflüsse von Händigkeit, Läsion, Performanz und Perfusion

58 Jutta L. Mueller
Mechanisms of auditory sentence comprehension in first and second language: An electrophysiological miniature grammar study

59 Franziska Biedermann
Auditorische Diskriminationsleistungen nach unilateralen Läsionen im Di- und Telenzephalon

60 Shirley-Ann Rüschemeyer
The Processing of Lexical Semantic and Syntactic Information in Spoken Sentences: Neuroimaging and Behavioral Studies of Native and Non-Native Speakers

61 Kerstin Leuckefeld
The Development of Argument Processing Mechanisms in German.
An Electrophysiological Investigation with School-Aged Children and Adults

62 Axel Christian Kühn
Bestimmung der Lateralisierung von Sprachprozessen unter besondere Berücksichtigung des temporalen Cortex, gemessen mit fMRT

63 Ann Pannekamp
Prosodische Informationsverarbeitung bei normalsprachlichem und deviantem Satzmaterial: Untersuchungen mit ereigniskorrelierten Hirnpotentialen

64 Jan Derrfuß
Functional specialization in the lateral frontal cortex: The role of the inferior frontal junction in cognitive control

65 Andrea Mona Philipp
The cognitive representation of tasks – Exploring the role of response modalities using the task-switching paradigm

66 Ulrike Toepel
Contrastive Topic and Focus Information in Discourse – Prosodic Realisation and Electrophysiological Brain Correlates

67 Karsten Müller
Die Anwendung von Spektral- und Waveletanalyse zur Untersuchung der Dynamik von BOLD-Zeitreihen verschiedener Hirnareale

68 Sonja A.Kotz
The role of the basal ganglia in auditory language processing: Evidence from ERP lesion studies and functional neuroimaging

69 Sonja Rossi
The role of proficiency in syntactic second language processing: Evidence from event-related brain potentials in German and Italian

70 Birte U. Forstmann
Behavioral and neural correlates of endogenous control processes in task switching

71 Silke Paulmann
Electrophysiological Evidence on the Processing of Emotional Prosody: Insights from Healthy and Patient Populations

72 Matthias L. Schroeter
Enlightening the Brain – Optical Imaging in Cognitive Neuroscience

73 Julia Reinholz
Interhemispheric interaction in object- and word-related visual areas

74 Evelyn C. Ferstl
The Functional Neuroanatomy of Text Comprehension

75 Miriam Gade
Aufgabeninhibition als Mechanismus der Konfliktreduktion zwischen Aufgabenrepräsentationen

76 Juliane Hofmann
Phonological, Morphological, and Semantic Aspects of Grammatical Gender Processing in German

77 Petra Augurzky
Attaching Relative Clauses in German – The Role of Implicit and Explicit Prosody in Sentence Processing

78 Uta Wolfensteller
Habituelle und arbiträre sensomotorische Verknüpfungen im lateralen prämotorischen Kortex des Menschen

79 Päivi Sivonen
Event-related brain activation in speech perception: From sensory to cognitive processes

80 Yun Nan
Music phrase structure perception: the neural basis, the effects of acculturation and musical training

81 Katrin Schulze
Neural Correlates of Working Memory for Verbal and Tonal Stimuli in Nonmusicians and Musicians With and Without Absolute Pitch

82 Korinna Eckstein
Interaktion von Syntax und Prosodie beim Sprachverstehen: Untersuchungen anhand ereigniskorrelierter Hirnpotentiale

83 Florian Th. Siebörger
Funktionelle Neuroanatomie des Textverstehens: Kohärenzbildung bei Witzen und anderen ungewöhnlichen Texten

84 Diana Böttger
Aktivität im Gamma-Frequenzbereich des EEG: Einfluss demographischer Faktoren und kognitiver Korrelate

85 Jörg Bahlmann
Neural correlates of the processing of linear and hierarchical artificial grammar rules: Electrophysiological and neuroimaging studies

86 Jan Zwickel
Specific Interference Effects Between Temporally Overlapping Action and Perception

87 Markus Ullsperger
Functional Neuroanatomy of Performance Monitoring: fMRI, ERP, and Patient Studies

88 Susanne Dietrich
Vom Brüllen zum Wort – MRT-Studien zur kognitiven Verarbeitung emotionaler Vokalisationen

89 Maren Schmidt-Kassow
What's Beat got to do with ist? The Influence of Meter on Syntactic Processing: ERP Evidence from Healthy and Patient populations

90 Monika Lück
Die Verarbeitung morphologisch komplexer Wörter bei Kindern im Schulalter: Neurophysiologische Korrelate der Entwicklung

91 Diana P. Szameitat
Perzeption und akustische Eigenschaften von Emotionen in menschlichem Lachen

92 Beate Sabisch
Mechanisms of auditory sentence comprehension in children with specific language impairment and children with developmental dyslexia: A neurophysiological investigation

93 Regine Oberecker
Grammatikverarbeitung im Kindesalter: EKP-Studien zum auditorischen Satzverstehen

94 Şükrü Barış Demiral
Incremental Argument Interpretation in Turkish Sentence Comprehension

95 Henning Holle
The Comprehension of Co-Speech Iconic Gestures: Behavioral, Electrophysiological and Neuroimaging Studies

96 Marcel Braß
Das inferior frontale Kreuzungsareal und seine Rolle bei der kognitiven Kontrolle unseres Verhaltens

97 Anna S. Hasting
Syntax in a blink: Early and automatic processing of syntactic rules as revealed by event-related brain potentials

98 Sebastian Jentschke
Neural Correlates of Processing Syntax in Music and Language – Influences of Development, Musical Training and Language Impairment

99 Amelie Mahlstedt
The Acquisition of Case marking Information as a Cue to Argument Interpretation in German
An Electrophysiological Investigation with Pre-school Children

100 Nikolaus Steinbeis
Investigating the meaning of music using EEG and fMRI

101 Tilmann A. Klein
Learning from errors: Genetic evidence for a central role of dopamine in human performance monitoring

102 Franziska Maria Korb
Die funktionelle Spezialisierung des lateralen präfrontalen Cortex: Untersuchungen mittels funktioneller Magnetresonanztomographie

103 Sonja Fleischhauer
Neuronale Verarbeitung emotionaler Prosodie und Syntax: die Rolle des verbalen Arbeitsgedächtnisses

104 Friederike Sophie Haupt
The component mapping problem: An investigation of grammatical function reanalysis in differing experimental contexts using eventrelated brain potentials

105 Jens Brauer
Functional development and structural maturation in the brain's neural network underlying language comprehension

106 Philipp Kanske
Exploring executive attention in emotion: ERP and fMRI evidence

107 Julia Grieser Painter
Music, meaning, and a semantic space for musical sounds

108 Daniela Sammler
The Neuroanatomical Overlap of Syntax Processing in Music and Language - Evidence from Lesion and Intracranial ERP Studies

109 Norbert Zmyj
Selective Imitation in One-Year-Olds: How a Model's Characteristics Influence Imitation

110 Thomas Fritz
Emotion investigated with music of variable valence – neurophysiology and cultural influence

111 Stefanie Regel
The comprehension of figurative language: Electrophysiological evidence on the processing of irony

112 Miriam Beisert
Transformation Rules in Tool Use

113 Veronika Krieghoff
Neural correlates of Intentional Actions

114 Andreja Bubić
Violation of expectations in sequence processing

115 Claudia Männel
Prosodic processing during language acquisition: Electrophysiological studies on intonational phrase processing

116 Konstanze Albrecht
Brain correlates of cognitive processes underlying intertemporal choice for self and other

117 Katrin Sakreida
Nicht-motorische Funktionen des prämotorischen Kortex:
Patientenstudien und funktionelle Bildgebung

118 Susann Wolff
The interplay of free word order and pro-drop in incremental sentence processing: Neurophysiological evidence from Japanese

119 Tim Raettig
The Cortical Infrastructure of Language Processing: Evidence from Functional and Anatomical Neuroimaging

120 Maria Golde
Premotor cortex contributions to abstract and action-related relational processing

121 Daniel S. Margulies
Resting-State Functional Connectivity fMRI: A new approach for assessing functional neuroanatomy in humans with applications to neuroanatomical, developmental and clinical questions

122 Franziska Süß
The interplay between attention and syntactic processes in the adult and developing brain: ERP evidences

123 Stefan Bode
From stimuli to motor responses: Decoding rules and decision mechanisms in the human brain

124 Christiane Diefenbach
Interactions between sentence comprehension and concurrent action:
The role of movement effects and timing

125 Moritz M. Daum
Mechanismen der frühkindlichen Entwicklung des Handlungsverständnisses

126 Jürgen Dukart
Contribution of FDG-PET and MRI to improve Understanding, Detection and Differentiation of Dementia

127 Kamal Kumar Choudhary
Incremental Argument Interpretation in a Split Ergative Language:
Neurophysiological Evidence from Hindi

128 Peggy Sparenberg
Filling the Gap: Temporal and Motor Aspects of the Mental Simulation of Occluded Actions

129 Luming Wang
The Influence of Animacy and Context on Word Order Processing: Neurophysiological Evidence from Mandarin Chinese

130 Barbara Ettrich
Beeinträchtigung frontomedianer Funktionen bei Schädel-Hirn-Trauma

131 Sandra Dietrich
Coordination of Unimanual Continuous Movements with External Events

132 R. Muralikrishnan
An Electrophysiological Investigation Of Tamil Dative-Subject Constructions

133 Christian Obermeier
Exploring the significance of task, timing and background noise on gesture-speech integration

134 Björn Herrmann
Grammar and perception: Dissociation of early auditory processes in the brain

135 Eugenia Solano-Castiella
In vivo anatomical segmentation of the human amygdala and parcellation of emotional processing

136 Marco Taubert
Plastizität im sensomotorischen System – Lerninduzierte Veränderungen in der Struktur und Funktion des menschlichen Gehirns

137 Patricia Garrido Vásquez
Emotion Processing in Parkinson's Disease:
The Role of Motor Symptom Asymmetry

138 Michael Schwartze
Adaptation to temporal structure

139 Christine S. Schipke
Processing Mechanisms of Argument Structure and Case-marking in Child Development: Neural Correlates and Behavioral Evidence

140 Sarah Jessen
Emotion Perception in the Multisensory Brain

141 Jane Neumann
Beyond activation detection: Advancing computational techniques for the analysis of functional MRI data

142 Franziska Knolle
Knowing what's next: The role of the cerebellum in generating predictions

143 Michael Skeide
Syntax and semantics networks in the developing brain

144 Sarah M. E. Gierhan
Brain networks for language
Anatomy and functional roles of neural pathways supporting language comprehension and repetition

145 Lars Meyer
The Working Memory of Argument-Verb Dependencies
Spatiotemporal Brain Dynamics during Sentence Processing

146 Benjamin Stahl
Treatment of Non-Fluent Aphasia through
Melody, Rhythm and Formulaic Language

147 Kathrin Rothermich
The rhythm's gonna get you: ERP and fMRI evidence on the interaction of metric and semantic processing

148 Julia Merrill
Song and Speech Perception – Evidence from fMRI, Lesion Studies and Musical Disorder

149 Klaus-Martin Krönke
Learning by Doing?
Gesture-Based Word-Learning and its Neural Correlates in Healthy Volunteers and Patients with Residual Aphasia

150 Lisa Joana Knoll
When the hedgehog kisses the frog
A functional and structural investigation of syntactic processing in the developing brain

151 Nadine Diersch
Action prediction in the aging mind

152 Thomas Dolk
A Referential Coding Account for the Social Simon Effect

153 Mareike Bacha-Trams
Neurotransmitter receptor distribution in Broca's area and the posterior superior temporal gyrus

154 Andrea Michaela Walter
The role of goal representations in action control

155 Anne Keitel
Action perception in development: The role of experience

156 Iris Nikola Knierim
Rules don't come easy: Investigating feedback-based learning of phonotactic rules in language.

157 Jan Schreiber
Plausibility Tracking: A method to evaluate anatomical connectivity and microstructural properties along fiber pathways

158 Katja Macher
Die Beteiligung des Cerebellums am verbalen Arbeitsgedächtnis

159 Julia Erb
The neural dynamics of perceptual adaptation to degraded speech

160 Philipp Kanske
Neural bases of emotional processing in affective disorders

161 David Moreno-Dominguez
Whole-brain cortical parcellation: A hierarchical method based on dMRI tractography

162 Maria Christine van der Steen
Temporal adaptation and anticipation mechanisms in sensorimotor synchronization

163 Antje Strauß
Neural oscillatory dynamics of spoken word recognition

164 Jonas Obleser
The brain dynamics of comprehending degraded speech

165 Corinna E. Bonhage
Memory and Prediction in Sentence Processing

S 2 Tania Singer, Bethany E. Kok, Boris Bornemann, Matthias Bolz, and Christina A. Bochow
The Resource Project
Background, Design, Samples, and Measurements

166 Anna Wilsch
Neural oscillations in auditory working memory

167 Dominique Goltz
Sustained Spatial Attention in Touch: Underlying Brain Areas and Their Interaction

168 Juliane Dinse
A Model-Based Cortical Parcellation Scheme for High-Resolution 7 Tesla MRI Data

169 Gesa Schaadt
Visual, Auditory, and Visual-Auditory Speech Processing in School Children with Writing Difficulties